

UNIVERZA V LJUBLJANI
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Bojan Bratuž

**DETEKCIJA IN KLASIFIKACIJA ZVOKOV BOBNOV
V REALNEM ČASU**

DIPLOMSKO DELO VISOKOŠOLSKEGA STROKOVNEGA ŠTUDIJA

Mentor:
doc. dr. Matija Marolt

Ljubljana, 2007

Kazalo

POVZETEK.....	1
1 UVOD	3
2 TEORETIČNO PREDZNANJE	5
2.1 ZVOK	5
2.1.1 Na kratko o nastanku zvoka	5
2.1.2 Jakost in glasnost zvoka	5
2.1.3 Frekvenca	7
2.1.4 Spekter zvoka	7
2.1.5 Uho.....	8
2.1.6 Obdelava informacij v možganih	9
2.2 SNEMANJE ZVOKA	9
2.2.1 Zgodovina.....	9
2.2.2 Digitalno snemanje	11
2.3 O BOBNIH	12
2.4 VOKALNO BOBNANJE (BEATBOXING).....	13
3 OSNOVE DIGITALNEGA PROCESIRANJA SIGNALOV	15
3.1 SIGNALI	15
3.2 VZORČENJE	16
3.3 KVANTIZACIJA	17
4 FOURIEROVA TRANSFORMACIJA	21
4.1 DISKRETNA FOURIEROVA TRANSFORMACIJA.....	22
4.2 HITRA FOURIEROVA TRANSFORMACIJA (FFT)	23
4.3 KRATKOČASOVNA FOURIEROVA TRANSFORMACIJA	23
5 DETEKCIJA BOBNOV	27
5.1 FILTRIRANJE	27
5.1.1 Splošno o filtrih	27
5.1.2 Nizkopasovni filter	28
5.1.3 Butterworthov filter	30
5.2 DETEKCIJA ZAČETKOV (ONSET DETECTION)	31
5.3 OKNJENJE (WINDOWING)	32
5.4 ANALIZA SPEKTRA.....	34
5.4.1 Pridobivanje značilk	35
5.4.2 Klasifikacija.....	36
6 IZVEDBA	37
6.1 ORODJA	37

6.2	METODA DELA.....	38
6.2.1	Zajem signala	38
6.2.2	Priprava oken analize.....	38
6.2.3	Računanje začetkov	39
6.2.4	Prepoznavanje udarca bobna	39
6.2.5	Končen rezultat in testiranje	40
6.3	OPIS DELOVANJA APLIKACIJE	40
7	SKLEP	43
	Slike	45
	Literatura.....	47

POVZETEK

V moji diplomski nalogi sem se ukvarjal z detekcijo osnovnih udarcev bobnov, natančneje vokalnega bobnanja. Razviti je bilo potrebno sistem, ki prepozna različne udarce bobnov v realnem času. V sistemu sem najprej uporabil metodo detekcije začetkov udarcev (angl. onset detection) v signalu. Za razlikovanje med udarci sem si v veliki meri pomagal s Fourierovo transformacijo, na podlagi katere dobimo frekvenčni spekter signala. Iz signala sem nato izračunal najpomembnejše značilke in na podlagi tega udarce razvrstil v posamezne razrede. Kot rezultat je nastala aplikacija, ki posamezne udarce v vhodnem signalu v realnem času zamenja z vnaprej shranjenimi zvoki bobnov ter jih predvaja. Aplikacijo lahko uporablja vsak, ki ima vsaj malo smisla za proizvodnjo udarcev z usti in je lahko v pomoč pri učenju razlik med osnovnimi udarci.

V uvodnem poglavju predstavljam temo, na katero se delo nanaša, omenjam tudi že raziskane teme na tem področju in jih na kratko komentiram. Na kratko opišem, kako sem se problema lotil in ga kasneje rešil.

V drugem poglavju podajam teoretične osnove, ki omogočajo bralcu brez predznanja vsaj okvirno razumevanje tematike. Najprej povem nekaj splošnih besed o zvoku in o digitalnem snemanju zvoka. Ker se moja tematika nanaša na transkripcijo ritma oziroma bobnov, natančneje vokalnega bobnanja, predstavim v poglavju še nekaj splošnih informacij o bobnih in vokalnem bobnanju.

V nadaljevanju prehajam v bolj tehnične opise, in sicer kaj signal sploh je in kako ga zajeti v digitalni obliki. To je tematika digitalnega procesiranja signalov, ki je glavni predmet obravnave pri računalniški obdelavi podatkov, s katero sem se srečal v začetni fazi razvoja aplikacije.

Naslednje, verjetno najpomembnejše teoretično poglavje, se nanaša na Fourierovo transformacijo, ki je eno izmed najpomembnejših matematičnih orodij za analizo linearno časovnih sistemov in predstavlja moderno orodje v današnjih telekomunikacijskih sistemih.

Peto poglavje je v celoti namenjeno razlagi detekcije bobnov. Opišem vse potrebne algoritme in metode, ki so potrebne za razvoj aplikacije, kot so filtriranje signalov, detekcija začetkov, pridobivanje značilk ter klasifikacija udarcev bobnov.

Sledi pregled metode in izvedbe dela, razlaga uporabljenih orodij, opis pristopa k delu, razlaga načrtovanja aplikacije po korakih, prikaz rezultatov dela ter opis delovanja aplikacije.

V sklepu naredim povzetek opravljenega dela, ocenim uspešnost ter zanesljivost aplikacije in na koncu podam možne izboljšave.

1 UVOD

Glasbeni inštrumenti so zelo zanimive umetnine. Že več tisoč let uporabljajo ljudje za proizvodnjo glasbil različne materiale, kot so les, živalska koža in kosti, zato se je na ta glasbila igralo na različne načine.

Prav tako ljudje že veliko let poznajo svojevrsten inštrument, to je človeški glas. Najprej so poznali le enoglasno petje v dokaj majhnih intervalih, skozi leta pa se je tudi to razvijalo. Danes zasledimo skoraj povsod večglasno petje, predvsem pri zborih in vokalnih skupinah. Ker človek ne pozna meja, se je odločil raziskovati samega sebe in iz tega se je razvila tehnika vokalnega bobnanja, ki je vedno bolj popularna. To je v bistvu posnemanje pravih bobnov. Vokalno bobnanje temelji na imitaciji zvoka, petju, proizvajanju ritmov in drugih spretnosti. Več informacij o tem dobi bralec na domači strani [4].

Na začetku se je seveda glasbo poslušalo samo v živo, dokler niso izumili tehnike snemanja zvoka. To je omogočalo poslušanje glasbe brez prisotnosti glasbenikov in glasbil. Snemanje glasbe sta najprej predstavila T. Edison in E. Berliner z izumom fonografa in gramofona [1]. Več let je bilo poznano le analogno snemanje, dokler niso v 70. letih 20. stoletja razvili digitalni zapis. To revolucionarno odkritje je toliko napredovalo, da je danes precej glasbe posneto v digitalni obliki, skoraj zagotovo pa bo v prihodnosti večina posnetkov na voljo digitalno.

Računalništvo pa je šlo še korak naprej. Razvili so metode zajemanja zvoka preko računalnika, ki kasneje omogoči ogromno obdelav in sprememb. Temu z eno besedo pravimo digitalno procesiranje signalov, ki je zelo natančno opisano v knjigah [2] in [3]. Postavi se vprašanje, kaj lahko s procesiranjem signalov dosežemo. Najbolj pogost odgovor na to je, da poizkušamo izvedeti čim več informacij o zvoku oziroma stvari, ki nas zanima. Temu pravimo pridobivanje informacij iz glasbe (angl. Music Information Retrieval-MIR), ki je dokaj popularno področje raziskav v zadnjem času. Uporablja se za avtomatsko razvrščanje zvočnih učinkov, razvrščanje zvrsti glasbe, prepoznavanje ritma, prepoznavanje višine tona, razpoznavanje določene melodije v večglasju, itd. Več o MIR si lahko bralec najde na [5], obstaja pa tudi vsakoletno tekmovanje na tem področju, kjer se oceni delo v preteklem letu. Seveda je uspešnost pridobivanja informacij različna. Precej je odvisno od tega, kakšen pristop k delu vzamemo oziroma katerih po navadi že razvitih algoritmov, se poslužujemo. Za delo sem si izbral prepoznavanje ritma ali natančnejše detekcijo vokalnega bobnanja, ker je tema še relativno neraziskana. Podrobnejši opis algoritma za prepoznavanje bobnov je v [6], ki ga bom v nadaljevanju predstavil tudi sam. Obstajajo tudi drugi, bolj zapleteni pristopi detekcije, kot je razvrščanje s pomočjo neodvisnega razvrščevalnega stroja (angl. Autonomous Classification Engine-ACE) [7], razvrščanje s pomočjo podpornih vektorjev [8], na voljo pa je tudi primerjava raznih metod klasifikacije, ki je opisana v članku [9].

Koraki v moji diplomski nalogi so bili: spoznavanje digitalnih signalov in zajemanja le-teh iz določenega vira, nato preučevanje o tem, kako signal posneti in predvajati. Spoznal sem še

osnove procesiranja signalov, da sem lahko uporabil metode filtriranja, računanja začetkov, okenskih funkcij ter Fourierove transformacije. Na koncu sem poizkušal še v določeni meri optimizirati sistem.

2 TEORETIČNO PREDZNANJE

Poglavje je namenjeno temu, da bralec pridobi najosnovnejše informacije, ki so potrebne za nadaljnje razumevanje. Najprej podajam nekaj informacij o tem, kaj pravzaprav je zvok in definiram nekaj njegovih lastnosti, za podrobnejši opis pa je na voljo članek [10]. Bralcu ponujam tudi krajši opis zgodovine snemanja zvoka, iz česar je razvidno, kako je bil včasih nek postopek zapleten in je danes trivialen in kakšne tehnologije snemanja se danes uporabljajo. Ker se tema mojega dela nanaša na prepoznavanje bobnov oziroma vokalnega bobnanja, ob koncu podam še nekaj o tem.

2.1 ZVOK

2.1.1 Na kratko o nastanku zvoka

Zvok nastaja zaradi spreminjanja tlaka v okolju. Zvok lahko opredelimo kot mehansko valovanje, ki se širi v dani snovi. Kadar razpravljamo o človeškem sluhu, je medij, ki omogoča slušne zaznave, zrak. Zvok pa lahko nastaja in se širi tudi v tekočini in trdnem mediju. V zraku se zvok pri normalnih pogojih širi s hitrostjo 340 m/s, v vodi 1500 m/s, v aluminiju pa okoli 5100m/s. Pri tem se zvočna energija prenaša z molekule na molekulo. V zvočnem valu si izmenjaje sledijo področja z višjim in nižjim tlakom – zgoščine in razredčine. Razlike med temi vrednostmi imenujemo zvočni tlak. Meriti ga je mogoče z mikrofonom, ki spremembe zvočnega tlaka prevede v električno napetost.

V ožjem pomenu definiramo kot *zvok* le zvočno valovanje s frekvencami v slišnem območju človeškega ušesa, to je med 20 Hz in 20.000 Hz (20 kHz). Zvok z nižjimi frekvencami imenujemo infrazvok, zvok z višjimi pa ultrazvok. Človeško uho je najobčutljivejše za zvok v frekvenčnem obsegu med 2.000 in 3.000 Hz.

Šum je zvok, v katerem so v večji ali manjši meri navzoče vse sinusne sestavine v nekem frekvenčnem intervalu; njegov spekter je zvezen. Aperiodični šum je pok.

Zvok lahko opredelimo z *amplitudo zvočnega tlaka* in *frekvenco* (f). Prva je povezana z glasnostjo, druga pa z višino tona.

2.1.2 Jakost in glasnost zvoka

Jakost zvoka je fizikalna količina, določena kot gostota energijskega toka zvočnega valovanja. Mednarodni sistem enot predpisuje za jakost zvoka izpeljano enoto W/m^2 . Ker se sluhovod akustično obnaša kot na eni strani odprta piščal, lahko pričakujemo, da je uho bolj občutljivo za tiste zvočne frekvence, pri katerih pride v sluhovodu do resonance. Za izražanje vrednosti zvočnega tlaka uporabljamo logaritemsko lestvico.

Stopnja zvočnega tlaka (angl. Sound Pressure Level-SPL) je podana z razmerjem

$$SPL = 20 * \log \frac{P}{P'}.$$
(2.1)

Referenčni tlak (P') pa znaša pri frekvenci 1000 Hz

$$2 * \frac{10^{-5} N}{m^2}.$$
(2.2)

Pri tem tlaku je pri človeku prag zaznave zvoka.

Glasnost zvoka je definirana tako, da povzročita dva tona z enako glasnostjo v ušesu približno enak občutek. Občutljivost ušesa na zvok je približno logaritemsko odvisna od jakosti zvoka, kar je upoštevano pri definiciji glasnosti:

$$\text{glasnost} = 10 * \log_{10} \frac{j}{j_0}.$$
(2.3)

Pri tem je j_0 najmanjša jakost zvoka, ki jo lahko z ušesom zaznamo. Zveza med jakostjo in glasnostjo zvoka je v splošnem odvisna od frekvence.

Povprečen prag jakosti leži blizu $10^{-12} W/m^2$, medtem ko je meja bolečine pri približno $1 W/m^2$.

Pogosto najdemo glasnost podano v *decibelih (dB)*. Tako definirano glasnost dobimo, če v enačbo (2.3) vstavimo $j_0 = 10^{-12} W/m^2$. To ustreza jakosti zvoka z glasnostjo nič decibelov pri frekvenci 1000 Hz. Glasnost, izražena v decibelih, ni odvisna od frekvence zvoka. Iz tega sledita dve pomembni dejstvi:

- Človekovo slušno območje sega od približno 0 dB, tj. stopnja, kjer je lahko slišen najmehkejši zvok, pa do nekje 125 dB, kjer je meja bolečine.
- Razlika med stopnjo amplitude, ki je manjša od 1dB, ustreza najmanjši razliki, ki je lahko slišna.

Glasnost zvoka lahko merimo tudi v fonih. Glasnost enega fona ustreza pri 1000 Hz (1 kHz) jakosti 1 dB.

Osnovna tabela glasnosti zvoka:

- 20 dB zvočni studio (zvočna izolacija)
- 50 dB urad
- 60 dB veleblagovnica
- 90 dB podzemna železnica

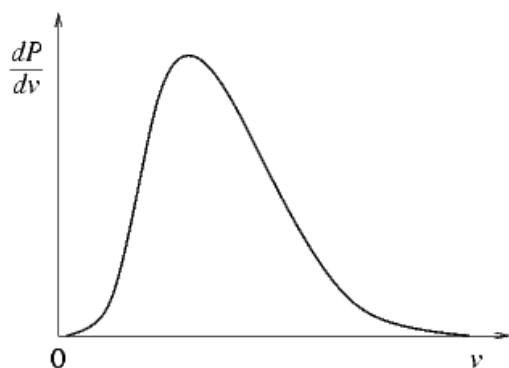
- 120 dB pnevmatično kladivo (prag bolečine)
- 160 dB reaktivno letalo
- 195 dB rakete pri vzletu

2.1.3 Frekvenca

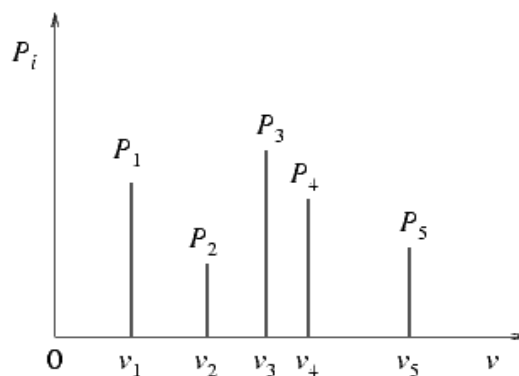
Drugi parameter, s katerim je določen zvok, je njegova *frekvenca*. To je fizikalna količina, določena kot število nihajev v časovni enoti. Obratna vrednost frekvence je nihajni čas $T = \frac{1}{f}$. Enota za merjenje frekvence je *hertz (Hz)*. Imenovana je v čast nemškemu fiziku Heinrichu Rudolfu Hertzu [10] in njegovim pomembnim prispevkom na področju elektrike in magnetizma. En hertz ustreza enemu nihaju na sekundo. Zvok, ki ga sestavlja ena sama frekvenca, imenujemo ton. Glasbeni zvoki vsebujejo več frekvenc: najnižjo ali osnovno ter višje frekvence. Slednje so mnogokratniki osnovne in jih imenujemo višji harmonični toni oziroma alikvoti. Zvok, v katerem so vse frekvence zastopane v približno enakem razmerju, imenujemo beli šum. Človeško uho zaznava zvočne vibracije v območju med 20 in 20.000 Hz. Slušni prag se spreminja s frekvenco zvoka. Občutljivost za zvok je največja pri frekvenci 1 kHz in znaša 4 dB ali 4 fone, v drugih frekvenčnih območjih pa se občutljivost manjša. Ta pojav lahko prikažemo s krivuljami enake glasnosti (izofonske krivulje). Pri glasnosti 120 do 130 dB prehaja zaznava zvoka v občutek bolečine. Takšni zvočni dražljaji običajno poškodujejo slušni organ. Zvoki, ki nastajajo med govorjenjem, so v območju največje slušne občutljivosti.

2.1.4 Spekter zvoka

Spekter zvoka je predstavitev zvoka oziroma določenega odseka v zvoku z jakostmi vibracij pri različnih frekvencah. Ponavadi ga prikažemo kot graf, definiramo pa kot porazdelitev energije zvoka ali zvočnega tlaka v določenem frekvenčnem območju. Kadar je porazdelitev diskretna, govorimo o *diskretnem spektru* (Slika 2.2), kadar je zvezna, pa o *zveznem spektru* (Slika 2.1). Pri zveznem spektru so na danem intervalu zastopane vse frekvence, pri diskretnem pa le določene frekvence oziroma spektralne komponente. Pri pretvorbi iz analognega v digitalni signal (več o tem v nadaljevanju), s čimer sem se ukvarjal v svojem delu, lahko zvok predstavimo le z diskretnim spektrom.



Slika 2.1: Zvezni spekter

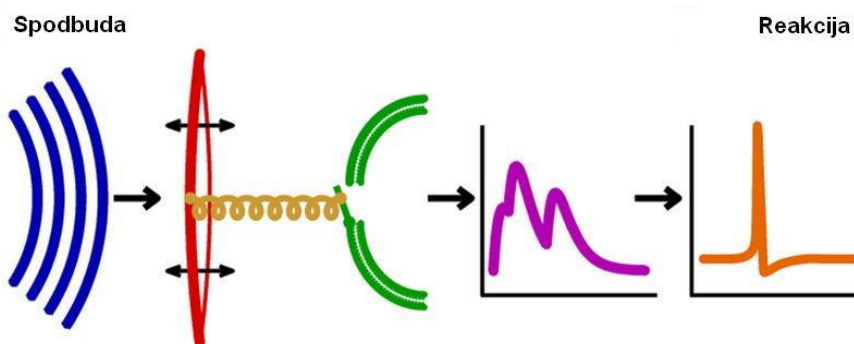


Slika 2.2: Diskretni spekter

2.1.5 Uho

Uho je slušni organ, ki pretvarja zvočne valove v živčne impulze. Sposobno je izločiti eno samo višino tona iz kompleksnega zvoka ali pa več tonov iz kompleksnih zvokov. Sestavljeno je iz treh glavnih delov: zunanjega, srednjega in notranjega ušesa. Na meji med zunanjim in srednjim ušesom je tanka opna, ki se imenuje bobnič. V votlini srednjega ušesa so zaporedoma povezane tri koščice: kladiyce, nakovalce in stremence, ki prevajajo tresljaje bobniča do ovalnega okenca na meji med srednjim in notranjim ušesom. Receptorji za zaznavanje zvoka so v polžu, delu notranjega ušesa. V notranjosti polža so tri vijačno zaviti zavinki (kanali). V preddvornem (vestibularnem) in bobničnem (timpanalnem) kanalu je perilimfa, tekočina z relativno visoko vsebnostjo natrijevih ionov. Med njima leži polžev (kohlearni) kanal, ki vsebuje endolimfo z relativno visoko vsebnostjo kalijevih ionov. Tresljaji bobniča se prenašajo po slušnih koščicah na ovalno okence in od tod po kanalih polža vse do okroglega okenca. Zanihanje kohlearnega kanala in bazilarne membrane povzroči ritmično upogibanje cilijarnih čutnih celic. Prednji del bazilarne membrane (bliže srednjemu ušesu) je najožji in najtanjši in zaniha pri visokih frekvencah. Del membrane ob polževi luknjici pri vrhu polža pa zaniha pri nizkih frekvencah. Kodiranje frekvence zvoka je torej neposredno odvisno od lege cilijarnih celic vzdolž polža.

Hkratno poslušanje z obema ušesoma nam omogoča, da določimo mesto izvora zvoka v prostoru, jakost in smer zvoka. Ker se zvok širi od izvora z enako hitrostjo, doseže obe ušesi istočasno le v primeru, če sta od zvočnega izvora enako oddaljeni. V drugih primerih dospe zvok prej do ušesa, ki je bližje izvoru, kar je osnova za določitev mesta zvočnega izvora v prostoru. Po tej teoriji se jakost zvoka kodira s frekvenco akcijskih potencialov iz ustreznih vlaken ter s številom vzdraženih vlaken, kar določa jakost zvoka. Smer zvoka pa določa zaostanek dražljaja med levim in desnim ušesom (pri nižjih frekvencah) ter razliko v jakosti med levim in desnim ušesom, ki bolj duši zvok na eni strani (pri višjih frekvencah).



Slika 2.3: Shematski prikaz sluha: zvočni valovi, bobnič, polž, avditorni receptorji, frekvenčni spekter slušnega odziva, živčni impulz

2.1.6 Obdelava informacij v možganih

V našem zunanjem in notranjem okolju neprestano potekajo energetske procesi, torej dogodki, ki omogočajo nastanek informacij. Senzorični sistemi imajo zmožnost sprejetja informacij, kodiranja in obdelave informacij v živčnih mrežah. Posamezno senzorično vlakno prevaja pretežno signale, ki jih povzroči določena frekvenca, čeprav se lahko vzdraži tudi ob drugih frekvencah. Krivulja vzdražnih pragov teh vlaken v odvisnosti od frekvence zvoka tvori njihovo resonančno krivuljo. V slušnem korteksu so celice organizirane v funkcionalne stolpiče. Nevroni posameznega stolpčiča imajo vse resonančne krivulje podobne. Slušne informacije v zvezi z govorom se takoj obdelajo v specializiranih območjih možganske skorje. Glavna območja možganske skorje s to vlogo so običajno v levi možganski polobli. To sta senčno (temporalno) govorno področje, imenovano tudi Wernickejevo polje, in čelno (frontalno) govorno področje ali Brockovo polje. Prvo sodeluje v slušnem procesu pri prepoznavanju besed, medtem ko je drugo bolj neposredno povezano z govorom.

2.2 SNEMANJE ZVOKA

2.2.1 Zgodovina

Zvočno snemanje ima bogato preteklost. Od trenutka, ko je bil zvok dejansko posnet in kasneje predvajan, pa vse do danes so se dogajale velike spremembe. Začetki segajo v leto 1870 z eksperimenti T. Edisona in E. Berlinerja [1]. Edison je leta 1877 izumil *fonograf* (Slika 2.4), pri katerem se je zvočno nihanje zapisalo z zarezi na valju, prevlečenim s kovinsko folijo, ki ga je bilo potrebno vrteti ročno. Za tem je Berliner leta 1887 patentiral *gramofon* (Slika 2.4). V primerjavi s fonografom je bila razlika v tem, da so bile zareze vtisnjene na ploski strani gramofonske plošče. Prihod električnega snemanja leta 1925 je drastično izboljšal kakovost snemalnega procesa na gramofonske plošče. Leta 1930 je tehnologija za reprodukcijo

hišnega zvoka doživela premiero. Sestavljala jo je kombinacija električno zapisanih gramofonskih plošč in posebej izdelanega ortofoničnega fonografa. To je bil peresno-vrezan akustičen fonograf, ki je uporabljal valovni vodnik in prepognjen rog, tako da je nudil razmeroma raven frekvenčni odziv.



Slika 2.4: Edisonov valjčni fonograf in gramofon na osnovi plošče

Za tem je prišlo magnetno snemanje, ki ga je v osnovi leta 1898 prikazal V. Poulsen s telegrafonom. Magnetno snemanje vsebuje rabo magnetiziranega medija, ki se premika z nespremenljivo hitrostjo čez snemalno glavo. Električni signal, ki je podoben zvoku, ki bo zapisan, povzroča primer magnetizacije, podoben signalu. Predvajalna glava lahko nato prebere spremembe v magnetnem polju traka in jih pretvori v električni signal. Z dodajanjem elektronskega ojačevanja se je leta 1920 telegrafon razvil v žične snemalnike, ki so bili med 1940. in 1950. letom priljubljeni za snemanje zvoka. Magnetno tračno snemanje, kot ga poznamo dandanes, je razvilo podjetje AEG leta 1930. To podjetje je med drugo svetovno vojno odkrilo tehniko poševnega tračnega snemanja (angl. biasing technique), kar je drastično izboljšalo kvaliteto zvoka in omogočilo razvoj snemalnikov do tehnične popolnosti.

Naslednji pomemben razvoj pri magnetnem traku je bilo večtračno snemanje, kjer je trak razdeljen na več prog, vzporednih druga z drugo. Prva faza večtračnega snemanja je bila stereo zvok, ki ima snemalno glavo razdeljeno na dve progi. To so najprej razvili nemški avdio inženirji leta 1943. Danes se v profesionalnih studiih uporablja 24-tračno snemanje tako, da se uporablja za en inštrument več prog.

Analogno zapisovanje na magnetni trak je imelo tudi slabost. To je šum, ponavadi imenovan sikanje (angl. hiss), ki ga povzroči končna velikost magnetnih delov na traku. Razmerje med signalom in šumom je večje pri visokih hitrostih in širokih trakovih, medtem ko se pri nižjih hitrostih in ožjih trakovih zmanjša. Leta 1963 je Phillips uvedel kompaktne avdio kasete, ki so v veliki meri pripomogle, da je format tračnega zapisovanja postal uporabnejši in kasneje začel dominirati na trgu, čeprav je po kvaliteti še vedno zaostajala za tehniko odprtega koluta. V 70. letih 20. stoletja sta koncept in trženje dodelanih analognih vezij postala ekonomično izvedljiva zaradi napredkov v elektroniki. To je pripeljalo do precejšnjih poskusov zmanjšanja sikanja s pomočjo uporabe različnih oblik stiskanja in raztezanja zvoka. Najbolj uspešni pri izdelavi teh sistemov so bili Dolby Laboratories.

2.2.2 Digitalno snemanje

V 80. letih je digitalno snemanje postopoma začelo nadomeščati analogno, čeprav tudi to ni popolnoma izginilo (nekateri profesionalni studii še vedno uporabljajo to tehnologijo). Digitalni magnetni trak se je na trgu pojavil leta 1986. Predstavlja standard za digitalni zapis in shranjevanje zvočnih signalov. Sloni na tehnologiji analognih video snemalnikov, ki za branje in zapisovanje na trak uporabljajo glave, nameščene na vrtljiv boben z osjo, ki je glede na trak zamaknjena za določen kot. Zaenkrat je najbolj pogosto uporabljen medij v studiih pred končno obdelavo signala (angl. mastering). Vse bolj pa ga zamenjujejo flash-kartice in trdi diski. Za arhiviranje se ni obnesel, saj ni bil dovolj zanesljiv. Po določenem času se namreč začne s traku izgubljati določeno število podatkov in sam sistem ni več sposoben obnovitve signala.

Snemanje je mogoče z vzorčno frekvenco 44,1 kHz, 48 kHz (EBU standard) in Long Play z vzorčno frekvenco 32 kHz. Vzorčenje je 16-bitno in signal ni stisnjen. V novem času pa so se že pojavili snemalniki, ki podpirajo tudi 20 ali celo 24-bitno vzorčenje. Dolžina posnetka na kaseti znaša lahko največ 125 minut. Slaba stran medija pa je, da je sposoben reproducirati posnetek samo v realnem času, ne pa kot novejši mediji, kjer je mogoče posnetek enostavno s pomočjo računalnika zajeti kot datoteko in je prenos podatkov nekajkrat hitrejši.

Na potrošniškem trgu sta trakove in gramofonske plošče zamenjala *zgoščenka* (angl. compact disc-CD) in v manjši meri tudi *minidisk*. Ta dva medija sta v celoti digitalna in za predvajanje potrebujejo zahtevno elektroniko. Prve zgoščeneke je začela izdelovati družba Philips in kasneje tudi Sony. Zgoščeneke so prišle na trg leta 1983 in se kmalu uveljavile med kupci. Leta 1985 je bil predstavljen prvi CD-ROM, ki je dobil svojo zapisljivo različico CD-R v zgodnjih 90. letih. Posnetki so zapisani z uporabo 16-bitnega PCM (pulzno kodna modulacija) kodiranja s 44,1 kHz vzorčenjem. Tehnologija glasbenih zgoščenek je bila kasneje prilagojena za shranjevanje podatkov v računalništvu. Za to obliko plošč je v uporabi oznaka CD-ROM.

Digitalni zvok lahko shranimo tudi na poljuben računalniški medij za shranjevanje. Z večanjem kapacitet trdih diskov in zmogljivosti procesorjev konec devetdesetih je postalo snemanje na trdi disk vse bolj priljubljeno. Obstajata dve obliki takega snemanja. Prva je ta, da s pomočjo namiznih ali prenosnih računalnikov kodiramo zvok prek vmesnikov v dvo ali več stezni digitalni zvok. Ti vmesniki so lahko notranje zvočne kartice ali zunanje naprave, ki jih lahko priklopimo prek USB ali Firewire vmesnika. Druga oblika pa je uporaba snemalnika, ki vsebuje analogno-digitalni pretvornik in obratno in enega ali dva prenosna trda diska. Taki snemalniki lahko vsebujejo 24 stez in so praktično enonamenski računalniki, ki so lahko povezani s standardnim računalnikom za obdelavo.

2.3 O BOBNIH

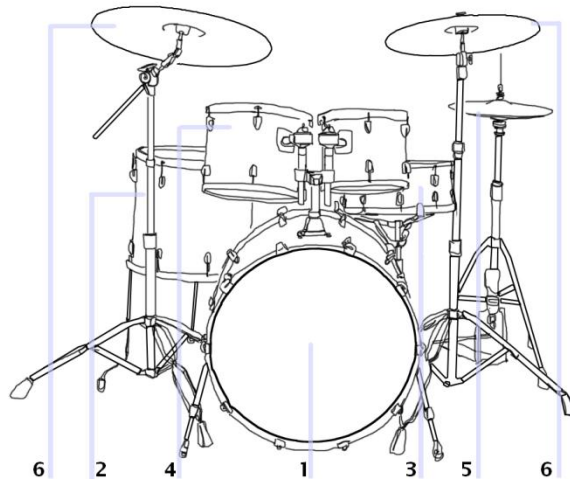
Boben je glasbeni inštrument, ki spada v družino tolkal, tehnično pa je razvrščen kot membranofon. To pomeni, da se zvok ustvari s pomočjo membrane (v obliki napete kože ali opne, ki je vpeta čez ustrezen nosilni okvir). Bobni so najstarejše in najbolj pogosto glasbilo na svetu, njihova oblika pa je ostala skoraj nespremenjena do danes.

Okvir bobna ima skoraj vedno okroglo odprtino, na katero je pritrjena opna, ostali del okvirja pa je lahko različnih oblik. Na zahodu zasledimo predvsem cilindrične oblike, čeprav so na primer timpani školjkaste oblike. Ostale oblike so lahko stožčaste (bongo bobni) ali v obliki keliha in druge. Boben cilindrične oblike je lahko odprt na eni strani (timpani) ali pa na obeh (dva opni). Cilinder je namenjen temu, da se lahko ustvari resonanca.

Današnji koncertni bobni so narejeni tako, da je opna pritrjena na obroč, nato pa ta obroč z enakomerno razporejenimi vijaki pritrjen na okvir. Vijakov je ponavadi od šest do deset, z njimi pa lahko nastavimo vpetost opne, kar je zelo pomembno za zvok bobna. Na zvok pa vplivajo še drugi dejavniki, kot so oblika bobna, velikost in debelina školjke, material, iz katerega je narejen okvir, vrsta opne, lega bobna in način, kako se nanj udari.

Običajno se na bobne igra z rokami ali pa z eno ali dvema palčkama. Pri mnogih tradicionalnih kulturah imajo bobni simbolično funkcijo in se ponavadi uporabljajo za verske obrede. V popularni glasbi si za bobne predstavljamo set bobnov (Slika 3.9) in bobnar je oseba, ki jih igra. Set bobnov je običajno zbirka bobnov in činel. Osnovni set vsebuje bas boben, snare boben in hihat činelo, s katerimi lahko zaigramo določen ritem. Bas boben (angl. kick) je največji boben in s tem tudi producira najnižje frekvence. V setu z ostalimi bobni se ga igra z nogo s pomočjo pedala, katerega glava udarja ob boben. Ta boben nam daje osnovni ritem v glasbi. Poznamo pa tudi orkestralni bas boben, na katerega se z eno roko igra, z drugo pa duši zvok. Drugi osnovni boben je snare boben, ki je cilindrične oblike in ima na vsaki strani eno opno. Okvir bobna je lahko narejen iz železa (ostrejši zvok) ali pa lesa (mehkejši zvok). Zvok producirajo s pomočjo različnih palčk, s katerimi se lahko udari na različne načine, na različnih mestih in z različno močjo udarca. Frekvenčno območje je pri tem bobnu nekoliko višje kot pri bas bobnu. Tretja iz osnovnega seta bobnov pa je hihat činela. Narejena je iz tankih kovinskih materialov, njena oblika pa je podobna krožniku. Za hihat potrebujemo dve manjši čineli (obrnjeni ena proti drugi) in stojalo, na katerega sta pritrjeni. Spodnja činela je fiksno pritrjena na stojalo, medtem ko je zgornja gibljiva in pritrjena na pedal in se jo lahko upravlja z nogami. Če sta čineli narazen, pravimo temu odprti hihat, če sta zaprti, pa zaprti oziroma samo hihat. Na njih lahko igramo s palčkami ali z nogami.

Naj omenim, da obstaja še veliko vrst drugih činel in bobnov, kot so tom bobni, crash činele, ride činele, tamburin, itd.



Slika 2.5: Set bobnov: 1-bas boben, 2-spodnji tom boben, 3-snare boben, 4-zgornji tom bobni, 5-hihat činela, 6-crash in ride činela

2.4 VOKALNO BOBNANJE (BEATBOXING)

Vokalno bobnanje je poleg udarcev s prsti za glasbenika najbolj intuitiven način izražanja ritma. Začetki segajo v zgodnja osemdeseta, ko se je uporabljalo kot osnova rap glasbenikom, nato pa je to postala oblika umetnosti. Najprej je vokalno bobnanje pomenilo nastajanje udarcev in ritmov s pomočjo človekovih ust, jezika, ustnic in grla. Vseeno pa to pritegne zraven še imitacije drugih zvokov, kot je na primer praskanje (angl. scratching). Dober vokalni bobnar ima ponavadi tudi široko območje drugih glasbenih spretnosti. Vokalno bobnanje temelji na imitaciji zvoka, petju, proizvajanju ritmov in drugih spretnosti. Največkrat ga povezujemo s hip-hop glasbo, srečamo pa ga tudi pri raznih vokalnih skupinah.

Tako kot pri pravih bobnih so tudi tukaj trije osnovni udarci, s katerimi proizvedemo večji del ritma. Nikjer ni predpisano, kako izvajati te udarce, vendar se je na določen način to mogoče naučiti. Tako lahko recimo bas boben proizvedemo, če izgovorimo 'p' ali 'b', za snare boben veljajo 'pf', 'psh', 'ps', 'k', za hihat pa 't', 'ts'. Teh treh osnovnih udarcev se ni težko naučiti. Težje je to spraviti v neko celoto, še težje pa nato obdržati konstanten ritem, na kar precej vpliva kondicija izvajalca. Obstajajo tudi podrobna navodila o vsem tem, ki so na vpogled na domači strani [4].

Vokalno bobnanje postaja vse bolj popularno in v tej smeri gre tudi napredek, saj se stalno srečujemo z novimi odkritji na robu človeških zmožnosti. Naj na koncu omenim še nekaj imen, ki so najbolj zaznamovala to umetnost. To so Keilla Kela, Shlomo, Rahzel, Biz Markie in drugi.

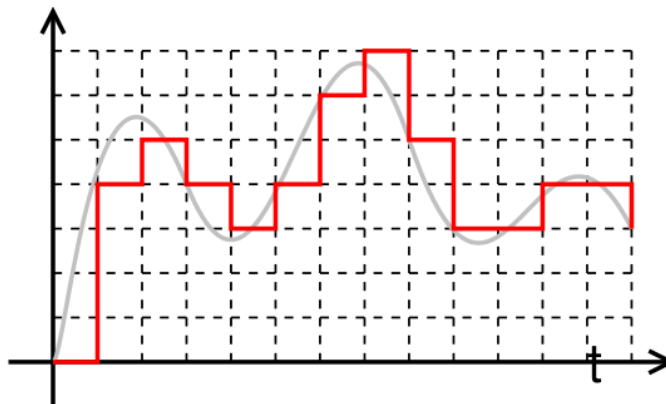
3 OSNOVE DIGITALNEGA PROCESIRANJA SIGNALOV

3.1 SIGNALI

Na vsakem koraku se srečujemo s signali [2]. Ti so naravnega ali umetnega izvora, njihov namen pa je prenos informacij. Generiranje signalov je povezano s sistemom, ki se odzove na nek dražljaj. Kombinacija dražljajev in sistema nam daje izvor signala. V nadaljevanju se bomo ukvarjali z enodimenzionalnimi signali, kot je govor ali zvok.

Če kakršenkoli signal pošljemo v sistem in gledamo izhod iz sistema, pravimo, da smo na izhodu dobili signal, ki je konvolucija vhodnega signala in prenosne funkcije sistema. Če imamo opravka s sistemom, ki je filter, pravimo, da smo na izhodu iz sistema dobili filtriran signal. Postopkom obdelave signalov pravimo s tujko tudi procesiranje signalov.

Ker so v vsakdanjem življenju signali večinoma analogni, moramo najprej ta signal vzorčiti in nato kvantizirati, šele nato ga lahko obdelujemo s postopki, ki so digitalne narave. Zato tudi pravimo tovrstnim postopkom digitalna obdelava signalov. Take obdelave omogočajo precej cenejšo obdelavo analognega signala. Postopek pretvorbe pa je tak, da signal spustimo skozi A/D (analogno/digitalni) pretvornik, nato lahko signal obdelujemo in če potrebujemo spet analogen signal, dodamo še D/A (digitalno/analogni) pretvornik. Prednosti digitalne obdelave je kar nekaj. Naj omenim večjo natančnost obdelave signalov, enostavnejše shranjevanje signalov, saj hranjenje na trdem disku ne vpliva na kvaliteto signala, pa tudi cena je v digitalni obliki mnogo manjša kot v analogni. Seveda pa ima digitalna obdelava tudi slabosti, kot so hitrost obdelave in hitrost vzorčenja signala. Analogne signale obravnavamo kot zvezne signale, digitalne pa kot diskretne.



Slika 3.1: Digitalni signal

3.2 VZORČENJE

V postopkih digitalne obdelave signalov se zatečemo k digitalnim postopkom, za katere velja, da moramo zvezne signale predstaviti s časovno diskretnimi signali.

To pomeni, da moramo analogne, časovno zvezne signale, najprej vzorčiti, nato pa jih kvantizirati oziroma jim določiti neko diskretno vrednost. Običajno diskretno predstavljen signal še kodiramo, kjer diskretne vrednosti predstavimo z binarnimi nizi.

Enakomerno vzorčenje je podano z izrazom

$$x(n) = x_n(t)|_{t=nT} = x_a(nT) \quad -\infty < n < \infty, \quad (3.1)$$

kjer je T interval vzorčenja in $x(n)$ vrednost analogno-zveznega signala v času nT .

Frekvenca vzorčenja je obratno sorazmerna intervalu vzorčenja in pove število vzorcev na enoto časa.

$$F_s = \frac{1}{T} \quad (3.2)$$

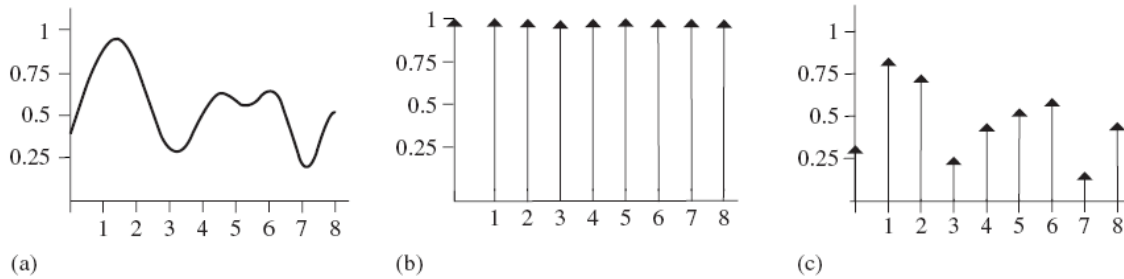
Pri vzorčenju zveznega signala prihaja do množenja (modulacije) med osnovnim zveznim signalom $x(t)$ in nizom Diracovih delta funkcij (vlak impulzov) $\delta(t)$ s periodo T_s . (Slika 3.2)

$$s(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT_s) \quad (3.3)$$

Vzorčeno zaporedje časovno zveznega signala $x(t)$ označimo z $x_s(t)$ in zapišemo

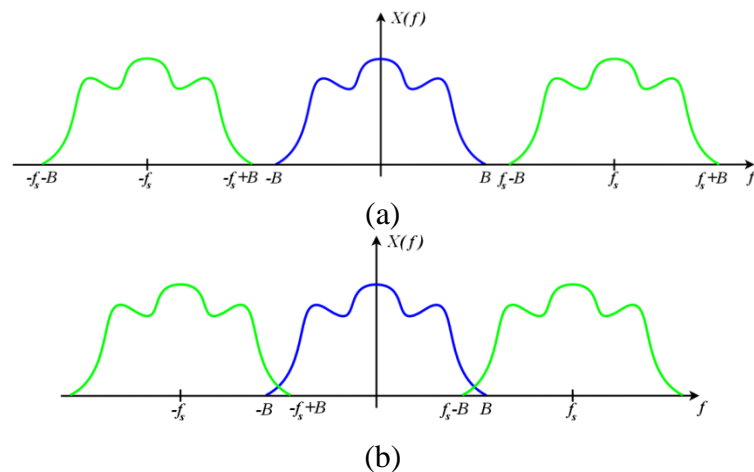
$$x_s(t) = x(t) * s(t) = \sum_{n=-\infty}^{+\infty} x(nT_s) * \delta(t - nT_s) \quad (3.4)$$

kar pomeni, da lahko predstavimo enačbo kot zaporedje enako oddaljenih impulzov, kjer je vsak utežen s funkcijsko vrednostjo funkcije $x(t)$ v trenutku nastopa impulza.



Slika 3.2: Zvezni signal (a), pomnožen z vlakom impulzov (b), nam vrne vzorčen signal (c)

Nyquistov zakon [11] pravi, da v primeru, da je dvakratnik frekvence vzorčenja F_s manjši od maksimalne frekvence, ki je prisotna v vzorčenem signalu, pride do prekrivanja (angl. aliasing-a). Potreben pogoj, da lahko iz diskretnega signala rekonstruiramo originalni signal je, da prekrivanja ni. Na sliki 3.3 je prikazan signal s prekrivanjem in brez prekrivanja.



Slika 3.3: (a) signal brez prekrivanja, (b) signal s prekrivanjem (aliasing)

3.3 KVANTIZACIJA

V praksi je kvantizacija zveznega signala v diskretni signal omejena s končno dolžino binarne besede, s katero predstavimo zvezni signal. Takšno pretvorbo realiziramo z analogno/digitalnim (A/D) pretvornikom. Osnovna naloga A/D je pretvorba zveznega amplitudnega signala v diskretno zaporedje digitalnih kod (besed). Ta pretvorba je sestavljena iz procesa kvantizacije in kodiranja. Kvantizator je nelinearen in nereverzibilen proces, ki priredi amplitudi določeno vrednost iz končne zaloge vrednosti.

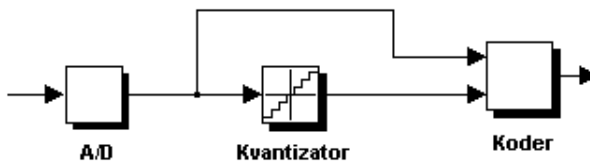
Dolžina besede A/D pretvornika je ponavadi 16 bitov in več. Zajeti vzorčni signal teoretično razdelimo od 2048 nivojev pri 10-bitnem pretvorniku in do 65536 (2^{16}) nivojev pri 16-bitnem pretvorniku.

Že samo zajemanje ni idealno, ker traja več časa, kot smo teoretično predpostavili, zato moramo pred A/D pretvornik postaviti zadrževalno vezje. Izhodni signal iz idealnega zadrževalnega vezja je tako

$$x_a(t) = h_0 * \sum_{n=-\infty}^{+\infty} x_a(nT) * \delta(t - nT). \quad (3.5)$$

Zadrževalna vezja so realizirana tako, da signal $x_a(t)$ vzorčijo v zelo kratkem času, nakar vezje zadrževalnika ohranja vrednost zajetega signala vse do naslednjega zajetega signala in to čim natančneje.

Idealna slika postopka kvantizacije in zadrževanja je sestavljena iz vzorčenja signala z zadrževalnikom in razmerjem med vhodnim in sledenim-zadržanim signalom. Model sistema vzorčenja in kvantizacije je prikazan spodaj (Slika 3.4)



Slika 3.4: Model sistema vzorčenja in kvantizacije

Kvantizator (A/D) je nelinearen sistem, ki pretvarja vhodne vzorce $x(n)$ v končni niz vnaprej predpisanih simbolov

$$\hat{x}[n] = Q(x[n]), \quad (3.6)$$

kjer je $\hat{x}[n]$ kvantizirana predstavitev $x(n)$. Govorimo o enolični kvantizaciji s konstantnim korakom.

V splošnem imamo 2^{B+1} kvantizacijskih nivojev. Razmerje med dolžino kodne besede in številom kvantizacijskih nivojev je odvisno od parametra, ki določa celotno področje A/D pretvornika. Tipični vrednosti A/D sta 10,5 in 1 volt. Nivo kvantizacije Δ pripada vrednosti bita z najmanjšo težo v kodni besedi. Za linearni kvantizator velja, da je

$$\Delta = \frac{x_{max} - x_{min}}{N - 1}. \quad (3.7)$$

Število kvantizacijskih nivojev N je določeno z izrazom $N = 2^b$, kjer je b število bitov na vzorec, Δ stopnica, x_{max} največja vrednost $x(n)$ in x_{min} najmanjša. Če povečamo število nivojev ($N - 1$), zmanjšamo velikost stopnice Δ . Tedaj se kvantizacijska napaka zmanjša in točnost A/D pretvornika poveča.

Ker območju Δ zveznega signala priredimo eno diskretno vrednost, pomeni, da je kvantizacija nereverzibilen proces, ki povzroča izgubo informacije. Velikost popačenja (izgube) je odvisna od števila bitov, ki jih uporabljamo pri kvantiziranju. To natančnost določimo glede na ceno A/D pretvornika in vzorčevalno frekvenco. Cena se večja z večanjem točnosti in frekvence.

Eden najbolj pogostih kvantizatorjev, ki se uporabljajo še danes, je pulzno – kodna modulacija (PCM). Bistvo te modulacije je, da je magnituda signala vzorčena v enakih intervalih in nato kvantizirana v zaporedje znakov v digitalni obliki. Takih kvantizatorjev se poslužujemo takrat, ko moramo dobljeni vzorčni signal še obdelati z na primer signalnim procesorjem.

4 FOURIEROVA TRANSFORMACIJA

Fourierova transformacija je eno izmed matematičnih orodij, uporabnih pri analizi linearno časovnih sistemov. Klasične Fourierove metode [12][3] kot sta Fourierova vrsta in Fourierov integral se uporabljata pri analizi časovno zveznih signalov in sistemov. To pomeni, da za signal $s(t)$ definiramo celoten interval $-\infty < t < \infty$, za vse t .

Predstavitev poljubnega periodičnega signala $x(t)$ lahko podamo z vsoto enosmerne komponente, osnovnega kosinusnega in sinusnega signala ter višjih harmonikov.

$$x(t) = a_0 + \sum_{k=-\infty}^{\infty} a_k \cos(w_k t) + b_k \sin(w_k t) \quad (4.1)$$

Enačba predstavlja zapis omenjenega periodičnega signala s Fourierovo vrsto. Konstantni del in vsak sinusoidni ali kosinusoidni izraz lahko kar imenujemo bazne ali osnovne funkcije. Fourierovi koeficienti so za vsa pozitivna števila definirani kot

$$w_k = k \frac{2\pi}{T} \quad (4.2)$$

$$a_k = \frac{2}{T} \int_{t_1}^{t_2} x(t) \cos(w_k t) dt \quad (4.3)$$

$$b_k = \frac{2}{T} \int_{t_1}^{t_2} x(t) \sin(w_k t) dt \quad (4.4)$$

$$a_0 = \frac{2}{T} \int_{t_1}^{t_2} x(t) dt \quad (4.5)$$

Koeficienti a_k so sodi Fourierovi koeficienti funkcije x , b_k lihi, w_k pa n -ti harmoniki funkcije. Harmonike lahko narišemo kot navpične črte na diagramu amplituda proti frekvenci.

Druga možna predstavitev signala je predstavitev v kompleksni eksponentni obliki, ki izgleda takole

$$x(t) = \sum_{k=-\infty}^{+\infty} c_k e^{i w_k t}, \quad (4.6)$$

kjer so posamezni koeficienti definirani kot

$$c_k = \frac{1}{T} \int_{t_1}^{t_2} x(t) e^{-i w_k t} dt \quad (4.7)$$

$$e^{-i w_k t} = \cos(w_k t) + i * \sin(w_k t) \quad (4.8)$$

4.1 DISKRETNA FOURIEROVA TRANSFORMACIJA

Do sedaj sem analiziral Fourierovo transformacijo zveznih periodičnih in aperiodičnih signalov, sedaj pa si oglejmo transformacijo končnega signalnega niza. To transformacijo imenujemo diskretna Fourierova transformacija – DFT.

DFT je dejansko transformacija niza N vzorcev in ni zvezna funkcija ter predstavlja enakomerno porazdeljene vzorce spektra $X(e^{i\omega})$ v točkah $\omega = \frac{2\pi k}{N}$, ki se ponavljajo z $2\pi r$. Kot bomo opazili, je Fourierova vrsta niza predstavljena z DFT končnega niza.

Integrali segajo pri diskretni verziji od 0 do T namesto od $-T$ do T kot pri zvezni. Če združimo a_k in b_k v eno kompleksno enačbo, dobimo

$$X_k = a_k - i b_k . \quad (4.9)$$

Če vpeljemo izraz (4.8) v enačbo (4.9), dobimo

$$X_k = \frac{1}{T} \int_0^T x(t) e^{-i w_k t} dt. \quad (4.10)$$

Gornji integral bi prišel v poštev pri zveznih funkcijah, za diskretne pa se integral spremeni v vrsto in tako dobimo

$$X_k = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{i \frac{2\pi k}{N} n}. \quad (4.11)$$

$x(n)$ predstavlja spekter periodičnega niza X_k in ga lahko izrazimo kot inverzno obliko Fourierove transformacije.

$$x(n) = \sum_{k=0}^{N-1} X_k e^{-i \frac{2\pi k}{N} n} \quad (4.12)$$

Časovna zahtevnost teh enačb je $O(N^2)$. Z določenimi izboljšavami pa lahko enačbo drugače zapišemo in tako lahko postane časovna kompleksnost $O(N \log_2 N)$. Implementacijo tega predstavlja hitra Fourierova transformacija (angl. Fast Fourier Transform-FFT).

4.2 HITRA FOURIEROVA TRANSFORMACIJA (FFT)

Postopke določitve DFT z manjšim številom operacij imenujemo postopki hitre Fourierove transformacije. Eden izmed takšnih postopkov je zasnovan na faktorizaciji, ki je odvisna od dolžine zaporedja N iz enačbe (4.11).

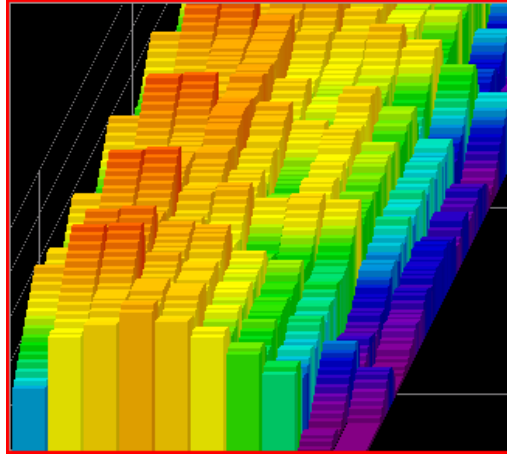
Ta algoritem sta predstavila Cooley in Tukey [13]. Pri njunem osnovnem postopku predpostavimo, da je N enak potenci števila 2, zaradi česar ga imenujemo algoritem osnove 2 (angl. radix-2 algorithm). Glede na način izvedene faktorizacije ločimo postopka določitve transformiranke z decimiranjem po času in decimiranjem po frekvenci. Eden od postopkov za določitev transformiranke je tudi ta, da je N enak potenci števila R . Te algoritme imenujemo algoritmi osnove R in izkaže se, da so najzanimivejši algoritmi osnove 4, 8 in 16.

Poznamo še algoritem deljene osnove (angl. split-radix), pri katerem je, tako kot pri algoritmu osnove 2, N enak potenci števila 2. Razlika je v tem, da faktoriziramo sodo indeksirane elemente kot pri postopku osnove 2, liho indeksirane elemente pa kot pri postopku osnove 4. Na ta način lahko definiramo zelo učinkovit postopek določitve DFT.

4.3 KRATKOČASOVNA FOURIEROVA TRANSFORMACIJA

Pri obravnavi problemov digitalnega procesiranja signalov, za katere je značilna predvsem časovna spremenljivost spektralne vsebine signala, se je pri njihovem reševanju izkazalo kot koristno kombinirati metode časovnega in frekvenčnega prostora. Če namreč takšne signale opišemo s Fourierovo transformacijo, imamo za ves dogodek na voljo eno samo Fourierovo transformiranko, zaradi česar je nemogoče ugotoviti dejanske značilnosti signalov. Kratkočasovna Fourierova transformacija (angl. Short Time Fourier Transform-STFT) temelji na ločenih transformacijah za vsak časovni trenutek. Če tako določimo Fourierove transformiranke za celotno trajanje signala in jih v grafični obliki predstavimo enega za drugim, dobimo sliko spreminjanja spektralne vsebine signala skozi čas.

V dvodimenzionalnem grafu predstavlja vodoravna os čas, navpična pa frekvenco. Amplitudo spektra prikazujemo z različnimi odtenki barv. Zgornjo teorijo lahko pri diskretnih signalih predstavimo s spektrogramom. Diagram lahko poleg dvodimenzionalne oblike predstavimo tudi v tridimenzionalni obliki, pri čemer je moč dobiti več informacij tudi o amplitudnem poteku posameznih frekvenčnih komponent. Tak graf prikazuje slika 4.1.



Slika 4.1: Tridimenzionalna predstavitev spektralnega spreminjanja

Časovno diskretna STFT zaporedja $x[m]$ je množica časovno diskretnih Fourierovih transformirank, ki ustrezajo različnim odsekom zaporedja $x[m]$. Časovni odsek za trenutek n definiramo z množenjem zaporedja $x[m]$ s pomaknjnim zaporedjem $w[n - m]$. Izraz za časovno diskretno STFT za trenutek n je tako definiran kot

$$STFT[m, \omega] = \sum_{m=-\infty}^{\infty} x[m]w[n - m]e^{-i\omega m}, \quad (4.13)$$

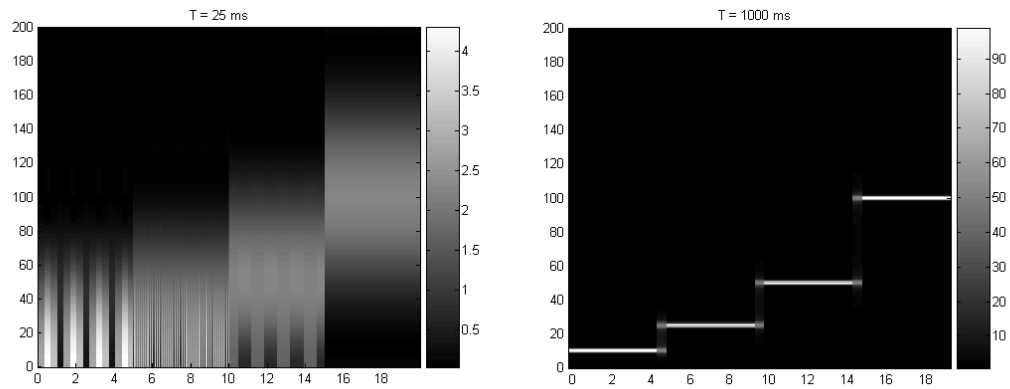
kjer zaporedje $w[m]$ predstavlja *okno analize*, $x[m]$ pa signal, ki ga transformiramo. $STFT[m, \omega]$ je kompleksna predstavitev faze in magnitude signala po času in frekvenci. Za funkcije $STFT[m + 1, \omega]$ pomaknemo okno analize za eno točko naprej po časovni osi, jo pomnožimo z $x[m]$ in izvedemo Fourierovo transformacijo nad odsekom. Če tako nadaljujemo, dobimo kot rezultat množico časovno diskretnih Fourierovih transformirank, ki skupaj predstavljajo časovno diskretno STFT in pokrivajo celoten signal. Okna analize se ponavadi prekrivajo. Pred izračunom Fourierove transformacije se okno pomnoži še z okensko funkcijo, več o tem pa v poglavju 5.3.

Magnituda pri STFT nam poda spektrogram signala, kjer je abscisna os časovna, ordinatna pa frekvenčna.

$$Spektrogram[m, \omega] = |STFT[m, \omega]|^2 \quad (4.14)$$

Ena od slabosti STFT je ta, da ima fiksno resolucijo. Širina okna nam pove, kako je signal predstavljen. Dolgo okno omogoča dobro frekvenčno in slabšo časovno razločljivost spektra, kratko okno pa omogoča dobro časovno in slabšo frekvenčno razločljivost. Naj za primer

navedem signal, ki vsebuje štiri frekvence (10, 25, 50, 100 Hz), ki se ne pojavljajo istočasno. Vzemimo še dve različni okni dolžin 25 ms in 1000 ms. Spektograma lepo prikazuje slika 4.2.



Slika 4.2: Spektograma dveh oken različnih dolžin

Na sliki z oknom velikosti 25 ms je lepo razvidno, kdaj pride do menjave signala, vendar je težko določiti njegove frekvence. Ravno obratno je pri oknu velikosti 1000 ms, kjer so jasno razvidne frekvence signalov, ni pa toliko jasno, kdaj te spremembe nastanejo.

5 DETEKCIJA BOBNOV

Poglavje je namenjeno dokončni implementaciji algoritmov in metod za detekcijo bobnov. Moje delo je bilo razdeljeno na dva dela, in sicer na detekcijo začetkov ter analizo in razvrščanje signala. Pri detekciji začetkov sem si pomagal z glajenjem signala s filtrom, zato najprej podam nekaj osnov filtriranja signalov in tudi bolj specifičnih filtrov. Temu sledi nadaljnja razlaga metode detekcije začetkov. Signal je bilo nato potrebno pretvoriti v frekvenčni spekter. Pretvorba je podrobno opisana v prejšnjem poglavju, dodam pa še nekaj besed o tem, kako odpraviti določene motnje oziroma izgube, ki so posledica neponavljajočega (diskretnega) signala. To opišem v poglavju oknjenje (angl. windowing). V nadaljevanju je potrebno signal še analizirati, kar pomeni, da poizkušamo iz njega pridobiti koristne informacije. Pri tem razložim metodi pridobivanja značilk in klasifikacije le-teh.

5.1 FILTRIRANJE

5.1.1 Splošno o filtrih

Filter je sistem, ki ločuje nekatere objekte ali parametre vhodnega signala od dela, ki ga dobimo na izhodu sistema. Najlažje si razložimo delovanje mehanskih filtrov, ki čistijo zrak ali pa ločujejo objekte od drobnejše vsebine.

Običajno obravnavamo linearne, časovno neodvisne filtre, kjer spekter vhodnega signala $x(\omega)$ spremenimo glede na prenosno funkcijo filtra $H(\omega)$. Tako dobimo odziv sistema

$$Y(\omega) = H(\omega) \cdot X(\omega). \quad (5.1)$$

$H(\omega)$ je prenosna funkcija sistema v frekvenčnem prostoru in jo tudi imenujemo utežna funkcija oziroma funkcija z oblikovanim spektrom. Filtre uporabljamo za odpravljanje neželenega šuma v signalu (neželenih frekvenc), za izbiro želenih frekvenc, za detekcijo signala pri npr. radarju, raznih sprejemnikih, za spektralno analizo, itd.

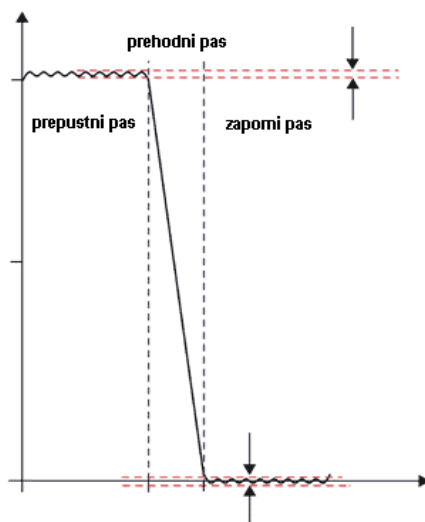
Poznamo digitalne in analogne linearne filtre. Čeprav so bili digitalni filtri sprva zgolj matematični nadomestek simulaciji obnašanja analognih filtrov na računalnikih, pa se je njihova uporaba s povečanjem zmogljivosti A/D pretvornikov in pojavom digitalnih signalnih procesorjev izjemno hitro razširila na praktično vsa področja elektrotehnike. K temu so v veliki meri pripomogli: možnost enostavnega spreminjanja lastnosti filtra brez posegov v njihovo zgradbo, velika zanesljivost delovanja, neobčutljivost na vplive okolice, natančnost izdelave, enostavnejša in predvsem cenejša velikoserijska proizvodnja digitalnih filtrov.

Glede na obliko sistemske funkcije $H(z)$ digitalnega filtra, ki ga navadno zapišemo v obliki

$$H(z) = \frac{b_0 + b_1 z^{-1} + \dots + b_M z^{-M}}{1 + a_1 z^{-1} + \dots + a_N z^{-N}}, \quad (5.2)$$

lahko filtre razdelimo v dve skupini: filtri z neskončnim impulznim odzivom (IIR filtri) in filtri s končnim impulznim odzivom (FIR filtri). Filtre pa lahko razdelimo tudi na več tipov, glede na frekvenčni odziv. Najbolj pogosti so *nizkopasovni* filter (angl. low-pass filter), ki prepušča nizke frekvence, *visokopasovni* filter (angl. high-pass filter), ki prepušča visoke frekvence, *pasovnoprepustni* filter (angl. band-pass filter), ki prepušča le določen pas frekvenc, *pasovnozaporni* filter (angl. band-stop filter), ki prepušča vse frekvence, razen določenega pasu, *vseprepustni* filter (angl. all-pass filter), ki prepušča vse frekvence, in drugi.

Običajen primer definiranja lastnosti filtra prikazuje slika 5.1, ki kaže karakteristiko nizkopasovnega filtra, kjer rdeči črtasti črti predstavljata dovoljeni toleranci prepustnega in zapornega pasu, navpični črtasti črti pa predstavljata mejni frekvenci prepustnega in zapornega pasu. Širina prepustnega pasu določa pasovno širino filtra. Prehodni pas je amplitudni potek odziva med prepustnim in zapornim pasom, ki enakomerno pojema ali narašča (visokopasovni filter). Krivulja prehodnega pasu je odvisna od načrtovanja filtra.



Slika 5.1: Amplitudne lastnosti nizkopasovnega digitalnega filtra

V svojem delu sem se ukvarjal z nizkopasovnim in visokopasovnim filtrom, a bom v nadaljevanju opisal samo prvega, saj sta si med seboj precej podobna.

5.1.2 Nizkopasovni filter

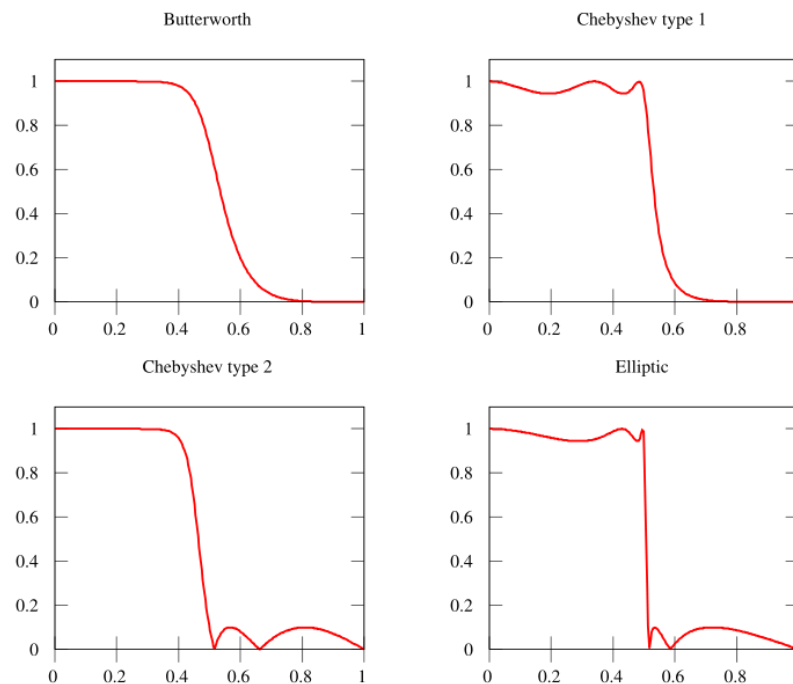
Kot sem že omenil, nizkopasovni filter prepušča nizke frekvence in je ravno obraten od visokopasovnega, s kombinacijo teh dveh pa lahko dobimo pasovnoprepustnega. Nizkopasovni filter se uporablja za različne namene, kot so filter sikanja (angl. hiss filter) v

avdio sistemih, za glajenje signala pri digitalnih algoritmih, za zvočne pregrade in drugi. Najlažje si ga predstavljamo, če smo v eni sobi in poslušamo zvok iz druge. Nizki toni so lahko slišni, medtem ko visokih skoraj ni.

Pri idealnem nizkopasovnem filtru bi bile vse frekvence po določeni meji prekinitve (angl. cutt-off frequency) popolnoma izločene. V realnosti to ne obstaja, zato v realnih aplikacijah aproksimiramo idealni filter z zamikom signala za majhno periodo. Temu zamiku pravimo fazni zamik. Kako hitro bo krivulja padala po meji prekinitve, določa *red filtra*. Najbolj običajni so filtri prvega in drugega reda, poznamo pa tudi višje.

Filter prvega reda ponavadi izgleda kot ravna črta do meje prekinitve in diagonalna po meji. Podobno izgleda krivulja tudi za filtre višjega reda, le da je diagonala vedno bolj strma, kar omogoča večje zmanjšanje visokih frekvenc. V splošnem je znano, da krivulja v enem koraku pade za določeno vrednost, ki je zmnožek reda filtra z -6 dB. Korak se zgodi takrat, ko se frekvenca podvoji.

Do sedaj sem opisoval osnovne lastnosti filtrov s standardno krivuljo, poznamo pa tudi prototipne filtre, ki se precej razlikujejo med seboj (Slika 5.2) in vsak ima svoje dobre in slabe lastnosti. Najpogostejši so Butterworthovi, Chebyshejevi, Eliptični, Besseljevi in Gaussovi tipi filtrov, obstajajo pa tudi drugi. Na vprašanje, kakšen od teh tipov je najuporabnejši, ni lahko odgovoriti. Če je pri načrtovanju pomembna zahteva po zvezni in linearni fazni karakteristiki, je najboljša izbira Butterworthov filter, če je zahteva po največjem dušenju v zapornem pasu, izberemo Eliptičen filter, če potrebujemo kompromis med omenjenima filtroma, bi bilo najbolje izbrati Chebyshev filter, itd.



Slika 5.2: Primerjava različnih prototipnih filtrov reda 5

5.1.3 Butterworthov filter

V svojem delu sem se odločil izbrati Butterworthov filter, zato naj podam še nekaj lastnosti tega filtra. Amplitudno karakteristiko $|H(\Omega)|$ Butterworthovih nizkopasovnih filtrov reda N , ki jih odlikuje predvsem konstantno ojačanje v celotnem prepustnem pasu brez kakršnekoli valovitosti, lahko zapišemo kot

$$|H(\Omega)| = \frac{1}{\sqrt{1 + \varepsilon^2 \left(\frac{\Omega}{\Omega_p}\right)^{2N}}}, \quad (5.3)$$

kjer je Ω_p mejna frekvenca prepustnega pasu, ε pa korelacijski faktor ojačanja prepustnega pasu. Če kot mejno frekvenco prepustnega pasu Ω_p definiramo kar mejno frekvenco Ω_c , pri kateri velja $|H(\Omega_p)| = -3dB$, dobimo za korelacijski faktor vrednost $\varepsilon = 1$, tako da se enačba (5.3) poenostavi v

$$|H(\Omega)| = \frac{1}{\sqrt{1 + \left(\frac{\Omega}{\Omega_c}\right)^{2N}}}. \quad (5.4)$$

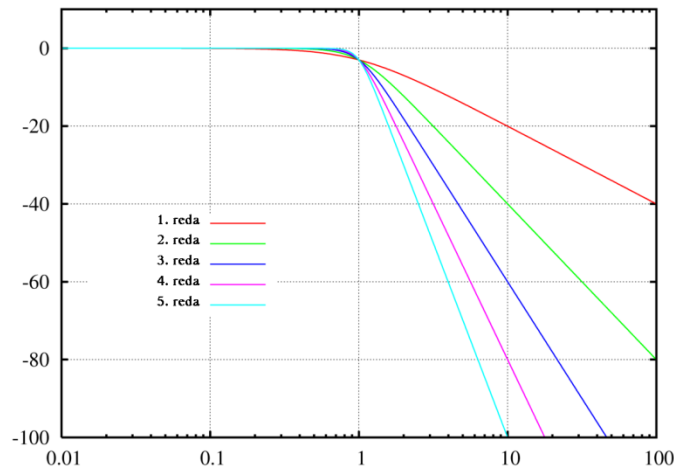
Red Butterworthovega filtra N s frekvencama prepustnega in zapornega pasu Ω_p in Ω_s ter valovitostima v prepustnem in zapornem pasu δ_1 in δ_2 je torej enak

$$N = \frac{\log \frac{\delta}{\varepsilon}}{2 \log \frac{\Omega_s}{\Omega_p}} \quad (5.5)$$

pri

$$\delta = \sqrt{\frac{1}{\delta_2^2 - 1}}. \quad (5.6)$$

Za primerjavo razlik med različnimi redi filtrov si oglejmo sliko 5.3, ki prikazuje krivulje različnih redov Butterworthovih nizkopasovnih filtrov.



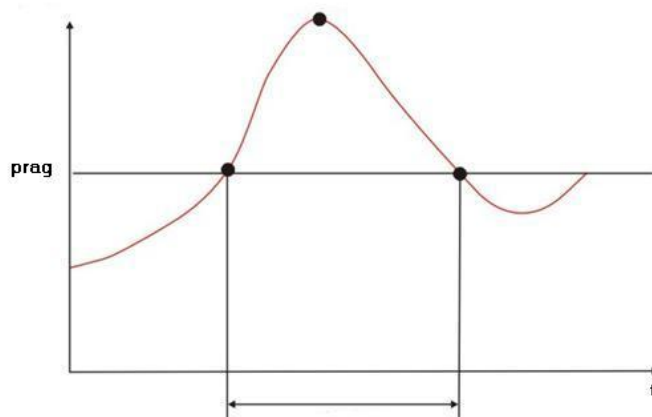
Slika 5.3: Krivulje amplitudnih karakteristik Butterworthovih nizkopasovnih filtrov reda 1 do 5

5.2 DETEKCIJA ZAČETKOV (ONSET DETECTION)

Pri analiziranju glasbe igra detekcija začetkov zelo pomembno vlogo, saj nam lahko pomaga pri določitvi tempa glasbe, izločevanje določene melodije, itd. Ker sem v svojem delu imel opravka z realnočasovnim računanjem, mi je detekcija začetkov pomagala pri določitvi začetka udarca, da ne bi bilo pred tem preveč nepotrebnega računanja in posledično zakasnitve. Dandanes obstaja že veliko učinkovitih algoritmov, ki se v glavnem razlikujejo v funkcijah detekcije. Te so lahko iskanje večjih sprememb v energiji signala, detekcija na podlagi faznega odklona, spektralnih okvirjev, podobnost med avdio signali, s pomočjo skritih markovih modelov, nevronskih mrež, itd. Več o tem si lahko bralec prebere v članku [14].

Obstaja pa tudi bistveno enostavnejša metoda, pri kateri lahko uporabimo nizkopasovni filter, opisan v poglavju 5.1.2. Omenil sem že, da lahko nizkopasovni filter uporabimo za glajenje signalov, kar nam omogoča lažje razpoznavanje značilnosti krivulje. Če so v signalu prisotne visoke frekvence, je samo nihanje signala hitrejše kot pri nizkih frekvencah. Ko pa ta signal spustimo skozi nizkopasovni filter, se nihanje precej zmanjša in če je mejna frekvenca filtra dovolj nizka, signal ne bo več nihal. Taki krivulji včasih rečemo tudi amplitudna ovojnica.

Tako lahko v vsakem oknu analize (glej poglavje 4.3) izračunamo absolutno vrednost signala, ki ga nato spustimo skozi prej omenjeni nizkopasovni filter. Preostane nam le še določitev praga (meje), ki izgleda kot vodoravna črta čez okno. Ko signal pride čez to mejo, takrat rečemo, da je to začetek. Za boljšo predstavbo si oglejmo sliko 5.4, na kateri nam prag določa začetek signala. Določitev višine praga je odvisna od namena uporabe.

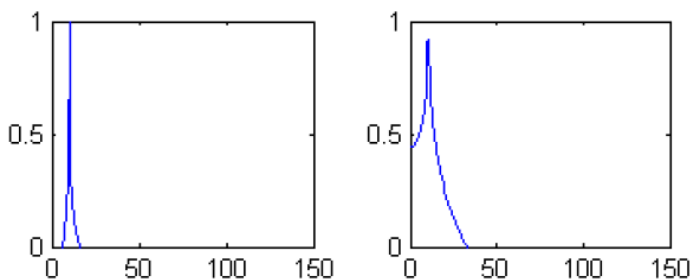


Slika 5.4: Detekcija začetka signala

Detekcija začetkov pri bobnih ni tako kompleksna, saj so udarci dovolj jasni in zato spremembe v krivulji velike. Tako je opisana metoda dala dovolj dobre rezultate in zato mi ni bilo potrebno posegati po kompleksnejših metodah, opisanih na začetku poglavja.

5.3 OKNJENJE (WINDOWING)

Kot sem že omenil v poglavju 4.2, je hitra Fourierova transformacija (FFT) namenjena pretvorbi signala iz časovnega prostora v frekvenčni prostor. FFT algoritem predpostavlja, da je signal neomejen (ponavljajoč), mi pa imamo opravka s časovno omejenim signalom. Večina realnih signalov ima ob koncu meritvenega časa prekinitev. Ker FFT algoritem predpostavlja, da je signal ponavljajoč, teh prekinitev ne bo predpostavljal. Druga slabost pa je ta, da pride do izgub v signalu. Vse to lahko povzroči, da je signal v frekvenčnem spektru popačen oziroma nerazpoznaven. Za primer si oglejmo sliko 5.5, kjer je na drugem grafu prikazan popačen signal.

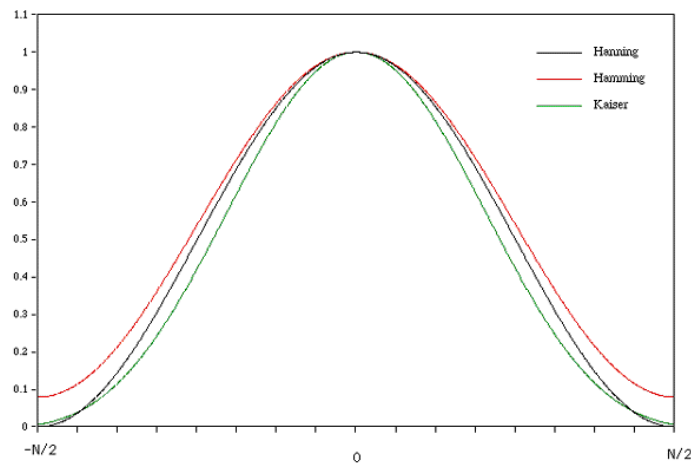


Slika 5.5: Primerjava ponavljajočega in neponavljajočega signala v frekvenčnem spektru s frekvenco 10 Hz.

Te slabosti pa lahko izboljšamo z oknjenjem. To je proces množenja časovnega okna signala z določeno funkcijo, ki na obeh straneh signal postopoma zmanjša do nič. Tej funkciji pravimo *okenska funkcija*. Oknjenje ponavadi izvajamo v časovni domeni, lahko pa jo tudi v frekvenčni, kar ne spremeni matematične metode.

Kakorkoli že, slabosti ne moremo v celoti odpraviti, jih pa lahko zmanjšamo. To je povzročilo, da danes obstaja veliko različnih okenskih funkcij. Vsako okno vpliva na spekter na drugačen način in vsako ima svoje dobre in slabe lastnosti, zato je izbira funkcije odvisna od namena uporabe. Najpogostejše funkcije so: Hanningova, Hammingova, Gaussova, Bartlettova, Blackmanova, Kaiserjeva, trikotna, pravokotna in druge. Za pregled razlik med temi funkcijami priporočam ogled članka [15].

V mojem primeru je bila najboljša izbira *Hammingove funkcije*, zato naj povem še nekaj o tem. Hammingova funkcija je v bistvu seštevek Hanningove in pravokotne funkcije. Krivulja funkcije se na začetku počasi vzpenja, na sredini doseže vrh in nato enako hitro pade. Slika 5.6 prikazuje krivuljo te funkcije v primerjavi s Hanningovo in Kaiserjevo funkcijo.



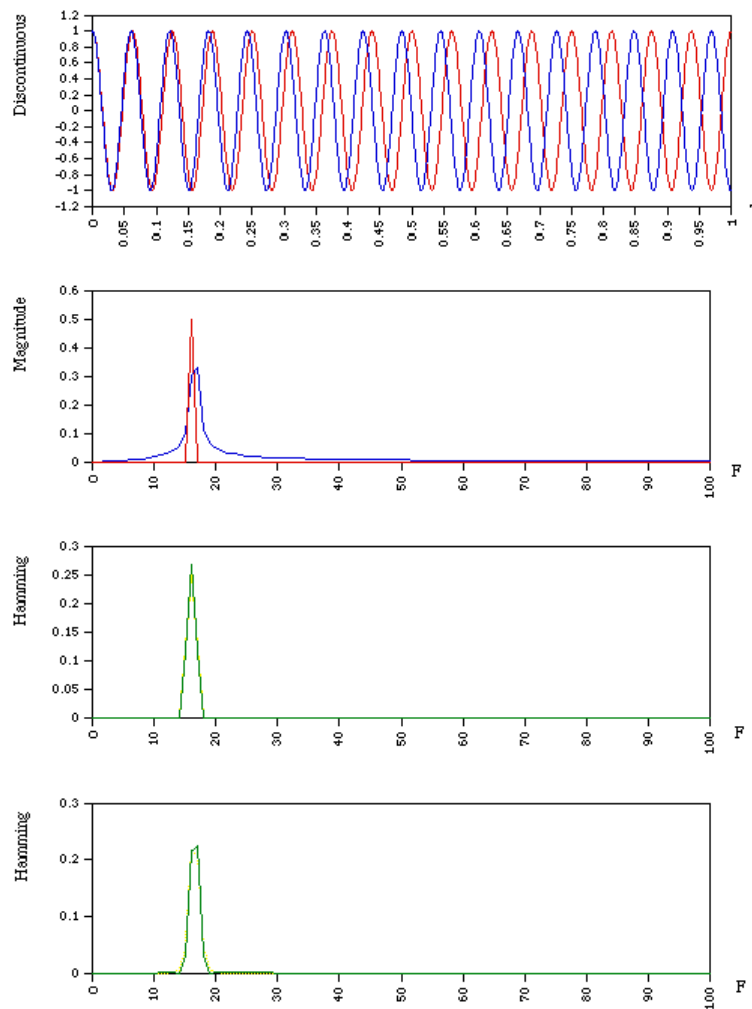
Slika 5.6: Primerjava treh podobnih okenskih funkcij

Formula za izračun Hammingove funkcije pa je

$$\omega_k = 0.54 - 0.46 \cdot \cos\left(2\pi \frac{k}{N-1}\right), \quad (5.7)$$

kjer N predstavlja širino okna, k pa cela števila v intervalu $0 \leq k \leq N-1$. To pomeni, da je k časovni interval skozi okno in ko je enak 0, funkcija ω_k doseže maksimum.

Naj na koncu tega poglavja še prikažem, kako izgleda signal, ko ga množimo s Hammingovo funkcijo. Slika 5.7 prikazuje funkcijo ponavljajočega (modra barva) in neponavljajočega (rdeča barva) signala v frekvenčnem prostoru.



Slika 5.7: Pretvorba ponavljajočega (modra) in neponavljajočega (rdeča) signala v frekvenčni prostor brez (druga slika) in s Hammingovo funkcijo (tretja, četrta slika)

5.4 ANALIZA SPEKTRA

Frekvenčni spekter signala nam ne pomaga veliko, če iz njega ne znamo razbrati informacij. Na svetu obstaja mnogo tehnik [1], kako iz zvoka razbrati informacijo. S pomočjo tega bi lahko preučevali učinek glasbe, strukturo elektronske glasbe, kar se lahko razširi v transkripcijo glasbe. Realnočasovna analiza spektra je kot nekakšno uho za glasbene sisteme, rezultat tega pa je najpogostejše prepoznavanje višine tona ali prepoznavanje ritma. Uporabniki pa danes niso zadovoljni le z analizo zvoka. Hočejo ga spremeniti in to nazaj

združiti z izvirnim zvokom. Zato danes obstajajo različni filtri in efekti, ki jih najraje uporabljajo glasbeniki.

Če hočemo iz zvoka pridobiti določeno informacijo, moramo najprej odstraniti nepomembne dele in poudariti pomembne. Delom z drugo besedo pravimo tudi značilke. Izbira značilk je zelo pomembna, saj mora biti smiselna glede na kontekst. Če hočemo razvrstiti zvoke v določene skupine, moramo najprej pridobiti značilke zvoka in jih na podlagi tega razvrstimo.

5.4.1 Pridobivanje značilk

S pomočjo značilk lahko zvoku določimo višino, glasnost, svetlost in pasovno širino skozi čas. Danes se navadno uporabljajo značilke na osnovi spektralnih lastnosti zvoka, kot so pogostost prečkanja ničle, težišče spektra, spektralni rolloff, RMS signala, cepstrum, MEL frekvenčni koeficienti in ostali. V svojem delu sem se odločil uporabiti le računanje težišča spektra, ker se je pri problemu to izkazalo kot dovolj dobra metoda.

5.4.1.1 Težišče spektra

Težišče spektra (angl. spectral centroid) se navadno povezuje s svetlostjo zvoka. Pri težišču se meri povprečno frekvenco, uteženo z amplitudo spektra. Standardna formula za izračun težišča spektra je

$$C = \frac{\sum C_i}{i}, \quad (5.8)$$

kjer je C_i težišče enega spektralnega okna, i pa število spektralnih oken zvoka. Spektralno okno je neko število vzorcev, ki je enako velikosti FFT. Težišče posameznega spektralnega okna pa je definirano kot

$$C = \frac{\sum f \cdot M[f]}{\sum M[f]}, \quad (5.9)$$

kjer je $M[f]$ magnituda pri frekvenci f . Višje vrednosti težišča ustrezajo bolj svetlim zvokom in imajo ponavadi večjo energijo v zgornjih predelih.

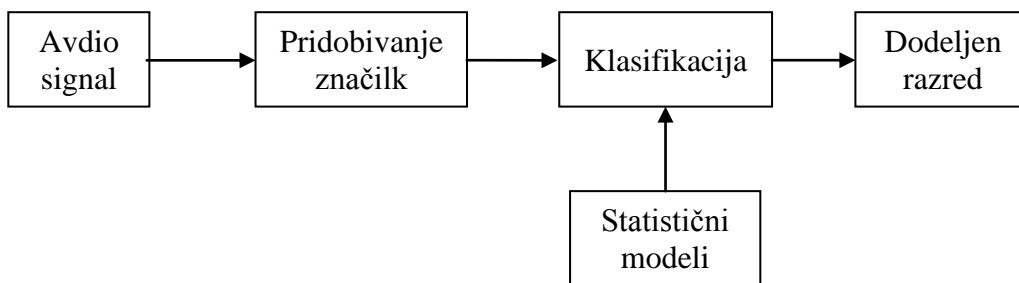
Naj za primer izračuna težišča vzamem tri frekvence (200, 400 in 900 Hz) z enakimi amplitudami. Izračunano težišče bo pri približno 750 Hz. Obstajajo tudi izračuni težišča v časovnem prostoru, vendar so zelo redki.

5.4.2 Klasifikacija

Ko imamo enkrat pridobljene značilke, je potrebno ta rezultat z določeno metodo ali več metodami klasificirati, kar pomeni, da jih razvrstimo v določene kategorije.

Vsak človek ves čas podzavestno razvršča avdio signale. Brez kakršnegakoli problema prepozna glas po telefonu ali pa razlikuje telefonski zvonec od hišnega. Problem pa nastane takrat, ko je zvok šibek, ima prisoten šum ali pa je podoben drugemu zvoku. V takem primeru je težko razvrstiti zvok, zato obstajajo računalniški sistemi za klasifikacijo signalov (angl. Audio Signal Classification-ASC) [16], ki potencialno zaznavajo veliko več kot človek in so mu tako lahko v veliko pomoč. Določen zvok lahko razvrstimo glede na zvrst glasbe, višino tona, ritem, lahko ugotavljamo ali gre za glasbo ali govor, itd. Z avtomatiziranjem teh lastnosti lahko dobimo sisteme za avtomatsko organizacijo podatkovnih baz, inteligentno analiziranje signalov, inteligentno kodiranje, avtomatsko izravnavo signala, avtomatsko kontrolo dinamike in druge.

Pri sistemih za avtomatsko klasifikacijo [17] je pomembno, da se odločimo za izbiro pravih značilk, ki opisujejo določen signal. Signal lahko nato primerjamo z zbirko vnaprej shranjenih modelov in se odločimo za tistega, ki se mu po značilnostih najbolj približa. Celoten potek avtomatskega razvrščanja je prikazan na sliki 5.8.



Slika 5.8: Potek avtomatskega razvrščanja zvoka

V svojem delu sem moral razvrstiti tri osnovne bobne (bas boben, snare boben in hihat činelo) po skupinah. Glede na težišče spektra, pridobljenega iz prejšnjega poglavja, sem se moral odločiti, kakšen boben udarec predstavlja. Na testnih primerih se je izkazalo, da ima bas boben težišče nekje od 100 Hz do 400 Hz, snare od 2500 Hz do 4000 Hz in hihat činela od 5500 Hz do 7500 Hz. Na podlagi teh ugotovitev sem sestavil tri razrede z oznakami 1, 2 in 3 (glede na boben) ter razred 0, ki je predstavljal slab udarec oziroma začetek. Na podlagi definiranih razredov sem kasneje za vsak nov udarec lahko enostavno ugotovil, v kateri razred spada.

6 IZVEDBA

Do sedaj sem opisoval teoretične osnove in različne metode in tehnike, ki sem jih uporabil v svojem delu, v tem poglavju pa bom predstavil dejansko implementacijo vsega tega in prikazal končni rezultat dela. Cilj mojega dela je bil prepoznavanje različnih udarcev bobnov iz človekovih ust (glej poglavje 2.4) preko mikrofona in te udarce zamenjati s pravimi bobni (glej poglavje 2.3). Pri tem sem se omejil le na tri osnovne tipe udarcev, to so bas boben, snare boben in hihat činela. Izziv v implementaciji mi je bil predvsem to, kako v realnem času čimprej zajeti signal, ga procesirati in predvajati končni signal. Kot opcijo v aplikaciji sem dodal še zajem signala iz wave datoteke, pri čemer čas izračuna ni tako bistvenega pomena, zato ta opcija omogoča še spremljanje različnih spektrov.

Celotno aplikacijo sem izdelal v okolju Visual Studio 2003 in jeziku C# in obsega zajem signala iz mikrofona oziroma vhodnega priključka na zvočni kartici ali pa iz zvočne (wave) datoteke, zatem zajem posameznih oken analize, računanja začetkov, filtriranja signalov, uporabo okenskih funkcij, računanje hitre Fourierove transformacije, računanje značilnk na podlagi dobljenega spektra, klasifikacijo in navsezadnje predvajanje zvoka. V pomoč so mi bile tudi zbirke iz okolja Measurement Studio, ki so olajšale prikazovanje grafov in implementacijo določenih algoritmov.

6.1 ORODJA

Microsoft Visual Studio 2003

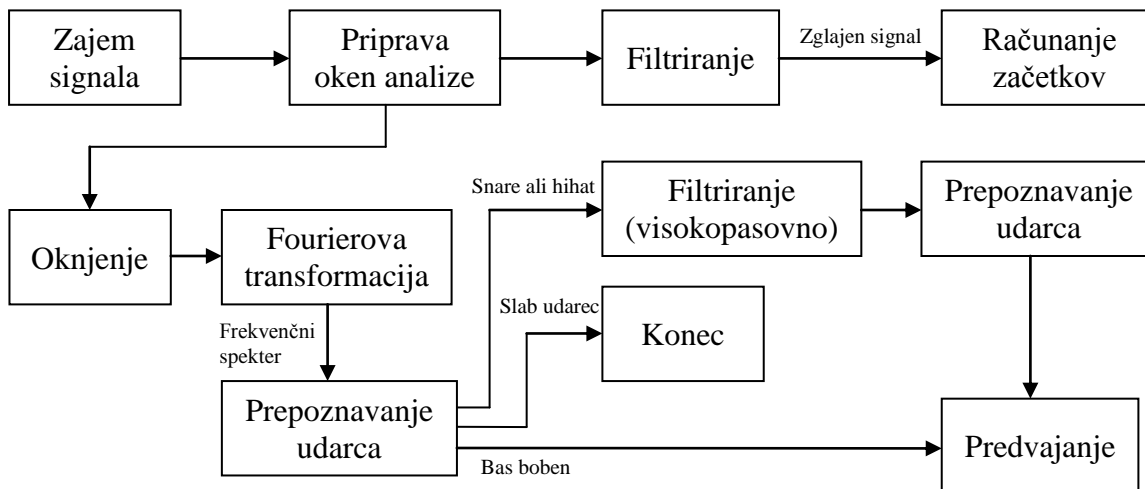
Visual Studio 2003 [19] je razvojno okolje, ki je namenjeno predvsem programerjem in omogoča izdelavo namiznih in spletnih aplikacij, spletnih strani in spletnih servisov. Zaganjajo se lahko na vseh platformah, ki podpirajo .NET Framework. Omogočajo kodiranje v več jezikih, kot so C#, Visual Basic, C++. Omogočajo tudi programiranje za mobilne aplikacije, v okolje pa se lahko vgradi mnogo zunanjih modulov oziroma aplikacij, ki nam razširijo paleto možnosti.

Measurement Studio for Visual C#.NET

Measurement Studio [18] je okolje, ki ga lahko integriramo v Visual Studio 2003 in je sestavljeno iz mnogih knjižnic in kontrol, namenjenim testiranju, opravljanju meritev in avtomatiziranje aplikacij. Okolje predstavlja orodja za pridobivanje, analizo predstavitev realnih podatkov.

6.2 METODA DELA

Delo je bilo sestavljeno iz več faz. Slika 6.1 v grobem prikazuje posamezne korake.



Slika 6.1: Metoda dela

6.2.1 Zajem signala

Da bi lahko karkoli razvijal, sem moral najprej zajeti nek signal (glej poglavje 3.1). V mojem primeru je bil to avdio signal. Razvil sem dve metodi, in sicer metodo zajemanja signala iz mikrofona oziroma vhoda zvočne kartice ter metodo zajemanja iz zvočne (wave) datoteke. Zvok sem zajel s frekvenco vzorčenja 44100 Hz, 16-bitno ter 2-kanalno. Zvok sem v obeh metodah zajemal v razmaku približno 50 ms, torej je bil pomnilnik za posamezno okno velikosti 2050 bytov.

V primeru zajemanja z mikrofonom sem enostavno zajel celotno okno v tem času in nadaljeval s procesiranjem, pri zajemu iz wave datoteke pa je bilo najprej treba odstraniti določene podatke v glavi datoteke in nato po vrsti prebrati vsebino. Več podatkov o wave formatu lahko bralec prebere v članku [20]. Od tu naprej so algoritmi za obe metodi enaki.

6.2.2 Priprava oken analize

Ni dovolj, da procesiramo okna analize (glej poglavje 4.3) eno za drugim, saj se lahko signal prične šele na koncu določenega okna in ta kratek odsek ni nujno pravilen kot eden od udarcev bobnov. Zaradi tega sem v aplikacijo dodal metodo, ki polovico predhodnega okna

prišteje k polovici trenutnega okna ter tako dobimo vmesno okno, ki je ravno tako dolžine 50 ms, ampak zamaknjeno za 25 ms nazaj. V združeni celoti to izgleda tako, da se okna križajo med seboj. S tem sem že v začetku povečal zanesljivost algoritma.

V eni proceduri imamo torej dve okni, eno za trenutni čas in eno za 25 ms pred tem. Za vsako od oken se nato v nadaljevanju procesiranje normalno izvaja naprej. Še pred nadaljevanjem je potrebno 'byte' tabelo pretvoriti v 'short', saj v prvi tabeli predstavljata po dva bita skupaj neko realno število v drugi tabeli, s katerim lahko obdelujemo podatke. 'Short' tabela pa ni nič drugega kot tabela celih števil. V jeziku C# se to enostavno izvede s pomočjo vgrajene metode *Buffer.BlockCopy*.

6.2.3 Računanje začetkov

Ali gre za začetek nekega signala, lahko izračunamo z metodo za detekcijo začetkov (glej poglavje 5.2). Tukaj sem predpostavil, da se znotraj intervala 75 ms lahko zgodi le en udarec, kar pomeni, da računam tri zaporedna okna analize in v tem intervalu moram ugotoviti, za kakšen udarec gre. Tu nastopi problem realnočasovne detekcije, saj imamo brez kakršnegakoli računanja 75 ms zakasnitve, kar pa za človeka še vedno ni moteče.

V metodi sem najprej vzel absolutno vrednost signala, tako da so bile vse vrednosti nad 0 ter nato z nizkopasovnim filtrom (glej poglavje 5.1.2) zgladil krivuljo signala. Dobljeni krivulji sem nastavil mejo, ko je bil signal dovolj močen, da bi lahko bil udarec. To mejo je mogoče med zagonom same aplikacije dinamično spreminjati, saj je moč signala odvisna od samega mikrofona oziroma izvora, pa tudi v nastavitvi glasnosti. Mejo si torej prilagodimo za lasten primer. Če metoda začetka ni zaznala, se je okno enostavno ignoriralo in zaznavalo naslednje. Ko pa sem naletel na začetek, sem od trenutnega okna naprej vzel še dve okni in v vsakem oknu poizkušal določiti udarec.

6.2.4 Prepoznavanje udarca bobna

Kot sem prej omenil, se je metoda za prepoznavanje udarca za vsak udarec izvedla točno trikrat. Ob vsakem klicu te metode se je sprožila druga metoda (metoda za detekcijo bas bobna), v kateri sem kot rezultat dobil odgovor, ali je udarec bas boben ali ni. Za tak postopek sem se odločil, ker je najlažje ločiti bas boben od ostalih, metodo pa bom opisal v nadaljevanju. Če se je izkazalo, da je udarec bas boben, sem nadaljnje računanje opustil ter v tabelo *DetectedBeats* zapisal 1, kar ustreza bas bobnu.

V nasprotnem primeru se je klicala druga metoda, ki je poizkušala prepoznati snare boben ali hihat činelo, lahko pa tudi nič od tega. Metoda je torej vrnila število 2, če je udarec ustrezal snare bobnu, število 3, če je udarec ustrezal hihat čineli in število 0, če ni šlo za udarec. Tudi te meritve sem zapisal v tabelo *DetectedBeats*. Ko se je tabela napolnila do svoje

meje (v našem primeru velikosti 3), sem iz tabele izbral število, ki se je najpogosteje ponovilo. To je bil končni rezultat klasifikacije (glej poglavje 5.4.2) in na podlagi tega se je na izhodu sprožil udarec, ki je ustrezal temu številu.

6.2.4.1 Metoda za detekcijo bas bobna

V tej metodi sem najprej nad celotnim signalom uporabil okensko funkcijo (glej poglavje 5.3) ter izračunal frekvenčni spekter s pomočjo hitre Fourierove transformacije (glej poglavje 4.2). Knjižica za izračun FFT spektra je del okolja Measurement Studio, ki kot rezultat vrne magnitudo in fazni odziv signala. Za nadaljnje računanje sem uporabil samo polovico okna, kajti druga polovica je zrcalna slika prve. Iz dobljenih magnitud sem nato izračunal težišče spektra (glej poglavje 5.4.1) in če je bil rezultat težišča nekje med 100 Hz in 800 Hz, potem je rezultat ustrezal bas bobnu.

6.2.4.2 Metoda za detekcijo ostalih udarcev

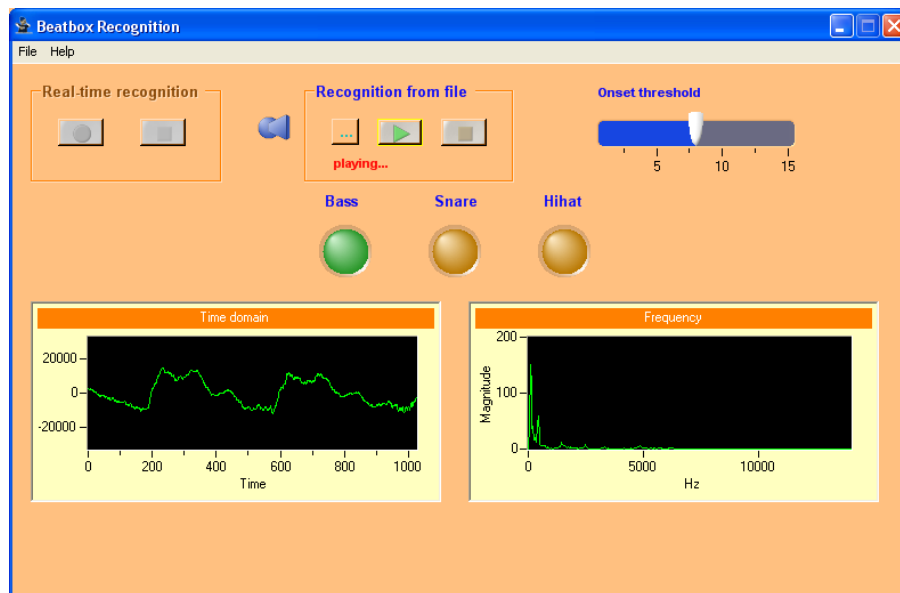
Tudi pri tej metodi sem uporabil okensko funkcijo ter izračunal frekvenčni spekter, le da sem pred tem uporabil še visokopasovni filter. S tem sem zmanjšal vpliv nizkih frekvenc in posledično lažje razlikoval med snare bobnom in hihat činelo. Za snare sem območje težišča vzel kar med 800 Hz in 5000 Hz, kar je bilo višje od 5000 Hz sem predpostavil za hihat činelo, kar pa je bilo nižje od 80 Hz, pa za slab signal oziroma udarec, torej razred 0.

6.2.5 Končen rezultat in testiranje

Aplikacija je na koncu predvajala le prave posnetke bobnov, ki so se sprožili glede na izvorni signal. Izvorni signal torej na izhodu ni bil predvajan. Kot opcijo sem v aplikacijo dodal še okni, ki sta v realnem času prikazovali časovni in frekvenčni spekter. Žal je izrisovanje teh dveh krivulj dodatno povečalo zakasnitev v realnem času, zato se grafa prikazujeta le za primer, ko je izvorni signal zvočna datoteka.

6.3 OPIS DELOVANJA APLIKACIJE

Ob zagonu datoteke 'Beatbox Recognition.exe', se nam odpre glavno okno aplikacije. Na voljo imamo meni z osnovnimi ukazi, glavni ukazi pa se nahajajo na namizju aplikacije (glej sliko 6.2).



Slika 6.2: Namizje aplikacije

Izbiramo lahko med realnočasovnim prepoznavanjem (Real-time recognition) in prepoznavanjem iz datoteke (Recognition from file). V prvem primeru se ob pritisku na gumb za snemanje začne prepoznavanje udarcev iz mikrofona ali kakšnega drugega vira, odvisno od nastavitev zvočne kartice. V drugem primeru lahko odpremo zvočno datoteko tipa *.wav in jo predvajamo. Grafa časovnega spektra (Time domain) in frekvenčnega spektra (Frequency) se izrisujeta le v primeru prepoznavanja iz datoteke. V obeh primerih pa se prižigajo zelene luči glede na to, katerega od bobnov je aplikacija zaznala. Na voljo imamo še drsnik (Onset threshold), s katerim določimo mejo zaznavanja udarcev. Potreben je za to, ker je jakost udarcev od ljudi do ljudi različna, nastavitve na zvočni kartici različne, pa tudi mikrofoni se razlikujejo med seboj.

Aplikacija je v celoti z izvorno kodo v jeziku C# priložena na CD-ju. Zraven je še inštalacijska datoteka, tako da se aplikacijo lahko namesti na kateremkoli računalniku, ki ima nameščen sistem Microsoft Windows XP ali 2000, ter Microsoft .NET Framework v1.1. Minimalne zahteve za delovanje aplikacije pa so procesor Intel Pentium III ali AMD Athlon, 512 MB pomnilnika ter zvočna kartica.

7 SKLEP

Cilj diplomske naloge je bil razvoj aplikacije za detekcijo vokalnega bobnanja v realnem času. Predstavil sem metode in tehnike, ki se uporabljajo za zajem signala, računanje začetkov, transformacije v frekvenčni prostor, filtriranje, analizo in klasifikacijo.

Na začetku je bila uspešnost prepoznavanja precej spreminjajoča, kajti težko je bilo razlikovati med snare bobnom in hihat činelo, ker imata oba prisotne nizke in visoke frekvence. To sem v večji meri odpravil z vpeljavo visokopasovnega filtra, zaradi katerega so bila težišča spektrov pri teh dveh udarcih višja in bolj oddaljena med seboj.

Aplikacijo sem optimiziral predvsem na lastno proizvedene udarce, pri katerih je bila uspešnost prepoznavanja zelo dobra. Nato sem opravil testiranja še na udarce drugih ljudi, kjer so se pojavile določene napake pri klasifikaciji. Težava je bila v tem, da vsak človek proizvede udarce z različno jakostjo in dolžino, drugačnimi frekvencami ter z ne vedno enako oddaljenostjo od mikrofona. Problem je nastal tudi pri prepoznavanju udarcev v realnem času. Sprva je večjo zakasnitev predstavljal prikaz grafov časovnega in frekvenčnega spektra, ki sem jih kasneje prikazoval le še pri prepoznavanju iz zvočne datoteke. Drugi problem pri zakasnitvi pa se je pojavil pri prenosu aplikacije na računalnik s slabšim procesorjem. Obdelava signala procesor precej obremeni, zato se aplikacija na starejših računalnikih slabo odziva.

Izboljšave bi bile smiselne predvsem pri izbiri značilk. V svojem delu sem uporabil le izračun težišča spektra, morda pa bi bilo smiselno uporabiti še RMS signala, ki nam določa glasnost signala, ali pa izračun sploščenosti spektra, ki nam pove, ali signal vsebuje veliko ekstremov ali ne. Izboljšave bi bile možne tudi pri klasifikaciji udarcev. Udarce sem klasificiral kar z določanjem mej pri izračunu težišča spektra. Na tem mestu bi se lahko odločil za klasifikacijo s pomočjo podpornih vektorjev [8], ki bolj učinkovito prepozna nezanestljive začetke, ali pa kakšno drugo metodo [9].

Slike

Slika 2.1: Zvezni spekter	8
Slika 2.2: Diskretni spekter	8
Slika 2.3: Shematski prikaz sluha: zvočni valovi, bobnič, polž, avditorni receptorji, frekvenčni spekter slušnega odziva, živčni impulz	9
Slika 2.4: Edisonov valjčni fonograf in gramofon na osnovi plošče	10
Slika 2.5: Set bobnov: 1-bas boben, 2-spodnji tom boben, 3-snare boben, 4-zgornji tom bobni, 5-hihat činela, 6-crash in ride činela	13
Slika 3.1: Digitalni signal	15
Slika 3.2: Zvezni signal (a), pomnožen z vlakom impulzov (b), nam vrne vzorčen signal (c)	17
Slika 3.3: (a) signal brez prekrivanja, (b) signal s prekrivanjem (aliasing)	17
Slika 3.4: Model sistema vzorčenja in kvantizacije	18
Slika 4.1: Tridimenzionalna predstavitev spektralnega spreminjanja	24
Slika 4.2: Spektograma dveh oken različnih dolžin	25
Slika 5.1: Amplitudne lastnosti nizkopasovnega digitalnega filtra	28
Slika 5.2: Primerjava različnih prototipnih filtrov reda 5	29
Slika 5.3: Krivulje amplitudnih karakteristik Butterworthovih nizkopasovnih filtrov reda 1 do 5	31
Slika 5.4: Detekcija začetka signala	32
Slika 5.5: Primerjava ponavljajočega in neponavljajočega signala v frekvenčnem spektu s frekvenco 10 Hz.	32
Slika 5.6: Primerjava treh podobnih okenskih funkcij	33
Slika 5.7: Pretvorba ponavljajočega (modra) in neponavljajočega (rdeča) signala v frekvenčni prostor brez (druga slika) in s Hammingovo funkcijo (tretja, četrt slika)	34
Slika 5.8: Potek avtomatskega razvrščanja zvoka	36
Slika 6.1: Metoda dela	38
Slika 6.2: Namizje aplikacije	41

Literatura

- [1] Roads C., The computer music tutorial, The MIT Press, London, England, 1996
- [2] Kačič Z., Digitalno procesiranje signalov, Tehniška fakulteta v Mariboru, 1994
- [3] Tasič J., Uvod v postopke digitalne obdelave signalov, Fakulteta za elektrotehniko, Ljubljana, 2002
- [4] Združenje vokalnih bobnarjev celega sveta. Dostopno na:
<http://www.humanbeatbox.com>
- [5] Portal iz področja pridobivanja informacij iz glasbe. Dostopno na: <http://www.music-ir.org/>
- [6] J. Sillanpää, Drum stroke recognition, Tampere university of technology, Finland. Dostopno na: <http://www.cs.tut.fi/sgn/arg/music/drums/raportti.ps>
- [7] E. Sinyor, C. McKay, R. Fiebrink, D. McEnnis, I. Fujinaga, Beatbox classification using ACE, McGill University, Quebec. Dostopno na:
<http://ismir2005.ismir.net/proceedings/2126.pdf>, 2005
- [8] S. Degroeve, K. Tanghe, B. De Baets, Support vector machines for bass and snare drum recognition, Ghent university, Belgium. Dostopno na:
http://www.ipem.ugent.be/mami/Public/Papers/GfKI2004_SVMForBassAndSnareDrumRecognition.pdf
- [9] P. Herrera, A. Yeterian, F. Gouyon, Automatic classification of drum sounds: A comparison of feature selection methods and classification techniques, Spain. Dostopno na: <http://www.iaa.upf.es/mtg/publications/ICMAI02-pherrera.pdf>
- [10] J. Lamancusa, Physics of sound, Penn State. Dostopno na:
http://www.mne.psu.edu/lamancusa/me458/5_physics.pdf, maj 2000
- [11] D. Lavry, Sampling theory for digital audio, Lavry engineering. Dostopno na:
http://www.lavryengineering.com/documents/Sampling_Theory.pdf, 2004
- [12] P. Bourke, Discrete Fourier transform, Fast Fourier transform. Dostopno na:
<http://local.wasp.uwa.edu.au/~pbourke/other/dft/>, jun. 1993
- [13] D. N. Rockmore, The FFT-an algorithm the whole family can use, Hanover. Dostopno na: <http://www.cs.dartmouth.edu/~rockmore/cse-fft.pdf>, oct. 1999
- [14] N. Collins, A comparison of sound onset detection algorithms with emphasis on psychoacoustically motivated detection functions, Spain. Dostopno na:
<http://www.cus.cam.ac.uk/~nc272/papers/pdfs/comparison.pdf>, 2005
- [15] FFT window functions, Bores signal processing. Dostopno na:
<http://www.bores.com/courses/advanced/windows/files/windows.pdf>
- [16] D. Gerhard, Audio signal classification: History and current techniques, university of Regina, Canada. Dostopno na:
<http://www2.cs.uregina.ca/~gerhard/publications/TRdbg-Audio.pdf>, 2003

- [17] G. Tzanetakis, Automatic musical genre classification of audio signals, Princeton. Dostopno na: <http://ismir2001.ismir.net/pdf/tzanetakis.pdf>
- [18] National instruments: Measurement studio for Visual C#.NET. Dostopno na: http://www.ni.com/mstudio/lrb_vc_sharp_net.htm
- [19] Microsoft: Visual studio 2003. Dostopno na: <http://msdn2.microsoft.com/en-us/vstudio/default.aspx>
- [20] S. Wilson, Wave PCM sound file format. Dostopno na: <http://ccrma.stanford.edu/courses/422/projects/WaveFormat/>, jan. 2003
- [21] Avtomatsko razvrščanje zvočnih učinkov, diplomsko delo, FRI, sept. 2005
- [22] Detekcija in klasifikacija šumnih dogodkov v glasbenih posnetkih, diplomsko delo, FRI, 2006

Zahvala

Zahvaljujem se mentorju doc. dr. Matiji Maroltu za koristne napotke in strokovno pomoč pri diplomski nalogi. Za podporo pri delu, pomoč pri oblikovanju in prevajanju sem zelo hvaležen puncu Karolini. Zahvala pa gre tudi prijatelju Roku Pahulje za dodatne napotke pri delu, Urški Povšič in Zdenki Plesničar za lektoriranje, ter moji družini za potrpežljivost.

Izjava o samostojnosti dela

Izjavljam, da sem diplomsko delo izdelal samostojno pod vodstvom mentorja doc. dr. Matije Marolta. Izkazano pomoč drugih sodelavcev sem v celoti navedel v zahvali.

Bojan Bratuž